

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

DIPLOMOVÁ PRÁCE



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY  
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

**ÚSTAV TELEKOMUNIKACÍ**

DEPARTMENT OF TELECOMMUNICATIONS

**EXTRAKCE PARAMETRŮ PRO VÝZKUM  
INTERPRETAČNÍHO VÝKONU**

EXTRACTION OF PARAMETERS FOR THE RESEARCH OF MUSIC PERFORMANCE

**DIPLOMOVÁ PRÁCE**

MASTER'S THESIS

**AUTOR PRÁCE**

AUTHOR

**Bc. Anna Laborová**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. Matěj Ištváněk**

**BRNO 2021**

# Diplomová práce

magisterský navazující studijní program **Audio inženýrství**  
specializace Zvuková produkce a nahrávání  
Ústav telekomunikací

**Studentka:** Bc. Anna Laborová

**ID:** 174461

**Ročník:** 2

**Akademický rok:** 2020/21

## NÁZEV TÉMATU:

### Extrakce parametrů pro výzkum interpretačního výkonu

#### POKYNY PRO VYPRACOVÁNÍ:

Prostudujte možnosti analýzy interpretačního výkonu v hudbě, možnosti detekce odchylek od notového zápisu a stylu hry využitím automatických metod pomocí číslicového zpracování zvukových signálů a metody parametrizace digitálního hudebního materiálu pro objektivizaci specifické analýzy hudebně-interpretačního výkonu. Pomocí vhodného programovacího prostředí proveďte analýzu hudebních nahrávek s využitím metod číslicového zpracování signálů.

Nejprve vytvořte návrh analýzy hudebních nahrávek z hlediska agogiky a porovnejte interpretaci stejného díla zahraničního různými interprety. Popište různé parametry, které mohou nejlépe definovat interpretační rozdíly, a vytvořte základní programové skripty pro detekci těchto parametrů. Následně proveďte akvizici databáze nahrávek, na které bude demonstrována parametrizace, a nakonec navrhnete vyhodnocovací systém. Práce bude obsahovat podrobný popis parametrů, funkční skripty pro jejich výpočet, kompletní databázi nahrávek pro demonstraci interpretačních rozdílů a statistické vyhodnocení dosažených výsledků podle dříve navrhnutého vyhodnocovacího systému.

#### DOPORUČENÁ LITERATURA:

[1] MÜLLER, Meinard. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Cham: Springer International Publishing, 2015. ISBN 978-3-319-21945-5.

[2] COOK, N.: Beyond the Score: Music as Performance. Oxford University Press 2013. ISBN 978-0-19-935743-7

**Termín zadání:** 1.2.2021

**Termín odevzdání:** 24.5.2021

**Vedoucí práce:** Ing. Matěj Ištváněk

**doc. Ing. Jiří Schimmel, Ph.D.**  
předseda rady studijního programu

#### UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

## ABSTRAKT

Rozdílné interpretace stejné skladby se mohou mezi sebou zásadně lišit. Nejen skladatel a noty definují posluchačův hudební zážitek, ale i charakteristická interpretace je integrální částí hudebního díla. Pro objektivní analýzu interpretačního výkonu jsou parametry rozděleny do čtyř tříd - z hlediska času (tempa), hlasitosti (dynamiky), témbra a výšky tónu. Každý parametr nebo jejich kombinace může ovlivnit výsledný unikátní charakter interpretace. Extrakce parametrů pro výzkum interpretačního výkonu je jedno z komplikovaných témat stojící na pomezí oblastí analýzy interpretačního výkonu a získávání informací z hudby. Předložená práce shrnuje poznatky a metody z obou oblastí. Pro výzkum byla vytvořena databáze, která obsahuje 31 nahrávek jedinečných interpretací 2. věty *Lento Smyčcového kvartetu č. 12 F dur* (1893) českého romantického skladatele Antonína Dvořáka (1841–1904).

## KLÍČOVÁ SLOVA

agogika, analýza interpretačního výkonu, artikulace, dynamika, extrakce parametrů, MIR, tempo, výzkum interpretace, získávání informací z hudby

## ABSTRACT

Different music performances of the same piece may significantly differ from each other. Not only the composer and the score defines the listener's music experience, but the music performance itself is an integral part of this experience. Four parameter classes can be used to describe a performance objectively: tempo and timing, loudness (dynamics), timbre, and pitch. Each of the individual parameters or their combination can generate a unique characteristic performance. The extraction of such objective parameters is one of the difficulties in the field of Music Performance Analysis and Music Information Retrieval. The submitted work summarizes knowledge and methods from both of the fields. The system is applied to extract data from 31 string quartet performances of 2. movement *Lento* of *String Quartet no. 12 F major* (1893) by czech romantic composer Antonín Dvořák (1841–1904).

## KEYWORDS

agogics, audio content analysis, articulation, dynamics, feature extraction, interpretation research, MIR, music performance analysis, music information retrieval, tempo

LABOROVÁ, Anna. *Extrakce parametrů pro výzkum interpretačního výkonu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2021, 63 s. Diplomová práce. Vedoucí práce: Ing. Matěj Ištváněk

## Prohlášení autora o původnosti díla

**Jméno a příjmení autora:** Bc. et Bc. Anna Laborová  
**VUT ID autora:** 174461  
**Typ práce:** Diplomová práce  
**Akademický rok:** 2020/21  
**Téma závěrečné práce:** Extrakce parametrů pro výzkum intepre-  
tačního výkonu

Prohlašuji, že svou závěrečnou práci jsem vypracovala samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno .....  
.....  
podpis autorky\*

---

\*Autor podepisuje pouze v tištěné verzi.

## PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu diplomové práce panu Ing. Matěji Ištvánkovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci. V neposlední řadě děkuji své rodině, že mi svou pomocí umožnila studium.

# Obsah

Úvod	11
<b>1 Analýza interpretačního výkonu</b>	<b>12</b>
1.1 Objektivizace analýzy interpretačního výkonu . . . . .	12
1.2 Předměty výzkumu interpretačního výkonu . . . . .	14
1.2.1 Tempo . . . . .	15
1.2.2 Dynamika . . . . .	16
1.2.3 Témbr . . . . .	16
1.2.4 Výška tónu a artikulace . . . . .	17
<b>2 Music Information Retrieval</b>	<b>19</b>
2.1 Music Performance Analysis . . . . .	20
<b>3 Parametrizace hudebního signálu</b>	<b>22</b>
3.1 Parametry popisující tempo . . . . .	22
3.2 Parametry popisující dynamiku . . . . .	23
3.3 Parametry popisující témbr . . . . .	24
3.4 Parametry popisující výšku tónu . . . . .	27
<b>4 Softwarové nástroje pro analýzu interpretačního výkonu</b>	<b>29</b>
4.1 Sonic Visualiser . . . . .	29
4.2 MIRtoolbox . . . . .	30
4.3 AMPACT . . . . .	30
4.4 librosa . . . . .	31
4.5 madmom . . . . .	31
<b>5 Databáze</b>	<b>32</b>
5.1 Výběr motivu . . . . .	32
5.2 Anotace . . . . .	32
<b>6 Synchronizace nahrávek</b>	<b>34</b>
6.1 Chroma reprezentace . . . . .	34
6.2 Dynamické borcení časové osy . . . . .	35
6.3 Vytvoření anotace . . . . .	37
<b>7 Návrh vyhodnocovacího systému</b>	<b>38</b>



<b>8 Analýza získaných parametrů</b>	<b>39</b>
8.1 Porovnání interpretací z hlediska agogiky . . . . .	40
8.2 Klasifikace . . . . .	47
<b>Závěr</b>	<b>52</b>
<b>Literatura</b>	<b>54</b>
<b>Seznam symbolů a zkratk</b>	<b>59</b>
<b>Seznam příloh</b>	<b>60</b>
<b>A Přiložené soubory</b>	<b>61</b>
<b>B Databáze nahrávek k analýze</b>	<b>62</b>
<b>C Notový zápis</b>	<b>63</b>

# Seznam obrázků

1.1	Komunikační hudební řetězec, od skladatele k posluchači. . . . .	14
3.1	Spektrogram nahrávky realizované souborem Smetanovo kvarteto (1959). . . . .	27
4.1	Uživatelské prostředí Sonic Visualiser. . . . .	30
5.1	Ukázka anotace úvodu 2. věty <i>Smyčcového kvartetu</i> č. 12 <i>F dur</i> v prostředí Sonic Visualiser. . . . .	33
6.1	Blokové schéma procesu synchronizace nahrávek a výběru stejného hudebního úseku. . . . .	34
6.2	Chromagram segmentu nahrávky realizované souborem Smetanovo kvarteto (1959). . . . .	35
6.3	Ukázka DTW matice $C$ s optimální cestou $P$ . . . . .	36
6.4	Ukázka synchronizace pozic taktů nahrávek Pavel Haas Quartet a Borusan Quartet. . . . .	37
7.1	Návrh vyhodnocovacího systému. . . . .	38
8.1	Závislost hodnot průměrného tempa podle roku vzniku nahrávky. . . . .	40
8.2	Notová transkripce taktu 1–6, 1. a 2. housle. . . . .	41
8.3	Srovnání délky taktů 1–6 všech interpretací z databáze. . . . .	41
8.4	Srovnání časového vývoje průměru českých a zahraničních nahrávek. . . . .	42
8.5	Škála barevnosti korelačního grafu (červená = silně pozitivní korelace, fialová = silně negativní korelace). . . . .	43
8.6	Korelace vývoje tempa nahrávek vybraných českých interpretů. . . . .	44
8.7	Porovnání vývoje RMS s průměrným průběhem BPM. . . . .	45
8.8	Míra <i>vibrata</i> u interpretace souborem Pavel Haas Quartet (2010). . . . .	46
8.9	Míra <i>vibrata</i> u interpretace souborem Musopen String Quartet (2012). . . . .	46
8.10	Shluky při závislosti RMS a tempa. . . . .	48
8.11	Shluky při závislosti MFCCs 6 a 7. . . . .	49
8.12	Shluky při závislosti MFCCs 10 a 11. . . . .	50
8.13	Shluky při závislosti tempa a spektrálního centroidu. . . . .	51
C.1	A. Dvořák - <i>Smyčcový kvartet</i> č. 12 „americký“, <i>F dur</i> . 2. věta Lento, intro. . . . .	63

# Seznam tabulek

1.1	Základní tempová označení . . . . .	16
1.2	Základní dynamická označení . . . . .	17
1.3	Přehled základních pojmů artikulace . . . . .	18
3.1	Tónové třídy . . . . .	28
8.1	Přehled extrahovaných parametrů . . . . .	39
8.2	Interpretace opakující se v hlavních shlucích . . . . .	51
B.1	Databáze nahrávek k analýze . . . . .	62

# Úvod

Interpretační výkon je důležitou dovedností všech hudebníků a interpretace je důležitou složkou hudebního díla samotného. Tradičně stojí v čele zájmu hudebních kritiků, ale méně často už muzikologů. Překvapivě je interpretace díla v rámci evropských dějin hudby a evropské estetické tradice druhořadou záležitostí. Ovšem s příchodem nahrávací technologie a záznamem hudebních provedení, a díky tomu zařazováním technických metod ke zkoumání hudby se otevírají nové otázky a možnosti zdokonalení výzkumu.

V souvislosti s interpretačním výkonem mluvíme o agogice. Agogika je zavedený termín, který popisuje odchylky hudebního provedení od notového předpisu. Především vyjadřuje modifikace tempa, které nejsou přesně zapsány notovým záznamem. Jev agogiky lze nalézt v každé hudební kultuře, ale nejvíce normován a sledován je v rámci evropské, resp. západní hudební oblasti, která se tradičně opírá o notový záznam. [1]

Příložená práce si klade dvě hlavní otázky:

- Jak extrahovat data, která splňují vysoké požadavky na spolehlivost a přesnost, tzn. že mají takový dostatečný význam pro vyvození obecných závěrů?
- Jak strukturovat a interpretovat extrahovanou informaci hudebně či muzikologicky smysluplným způsobem?

Tato práce se věnuje oblasti DSP (Digital Signal Processing – číslicové zpracování signálů), konkrétně podoblasti MIR (Music Information Retrieval – získávání informací o hudbě). Dává si za cíl zmapovat problematiku analýzy interpretačního výkonu a prozkoumat možnosti extrakce hudebních parametrů za účelem posuzování rozdílů mezi odlišnými interpretacemi stejné skladby. V úvodu je nutné upřesnit, že se práce zaměřuje na západní hudební kulturu, která používá notační systém založený na temperovaném ladění.

Teoretická část se v kapitole 1 věnuje úvodu do problematiky tradičních metod analýzy interpretačního výkonu a jeho charakteristice z muzikologického hlediska. Kapitola 2 mapuje stručnou historii a oblasti výzkumu disciplíny získávání informací o hudbě. Následující kapitoly 3 a 4 se zabývají popisem vybraných parametrů a popisem známých nástrojů pro jejich extrakci.

Praktická část práce se nejprve věnuje v kapitole 5 popisu vytvořené databáze. Kapitola 6 vysvětluje postup při synchronizaci nahrávek a vytvoření anotace u nahrávek z databáze podle anotace referenční. Extrakci parametrů a následné analýze z hlediska agogiky se věnuje kapitola 8.

Závěr poskytuje prostor diskuzi nad dosaženými výsledky diplomové práce.

# 1 Analýza interpretačního výkonu

Interpretace byla (a zčásti ještě je) v průřezu dějin hudby západní kultury považována za méně významnou část hudebního díla. Interpretační výkon může být vnímán pouze jako reprodukce, převod notového záznamu do zvukové podoby. Uznávání hudební teoretici, jako Johannes Brahms (1833–1897) nebo Heinrich Schenker (1868–1935), označovali interpretaci za „potlačování vlastní individuality“ nebo naopak za „zbytečný individuální přínos“. Proč tomu ale tak bylo či je, když bez interpretace je hudební dílo pouhým záznamem na papíře a když právě skrze interpreta hudební dílo ožívá a dosahuje kompletnosti, zůstává s otazníkem. [30]

Zásadním mezníkem ve vývoji umění ve 20. století je nesmírný rozvoj technických prostředků a masových médií. Zdokonalující se technologie zvukového záznamu, a také její zpřístupnění rostoucímu množství hudebníků a posluchačů (díky zvyšujícímu se počtu nahrávacích studií, rozvoji rozhlasu, televize a filmu), hrála klíčovou roli při utváření nových směrů a žánrů. Dala taktéž směr moderní interpretační tradici a v moderní muzikologii, především zahraniční, je oblast zkoumání metod analýzy interpretačního výkonu aktuálním trendem.

Specifickým podoborem muzikologie je *computational musicology*<sup>1</sup>. Jedná se o oblast výzkumu na hranici mezi muzikologií a počítačovou vědou. Zahrnuje veškeré disciplíny, které využívají výpočetní metody jako nástroje ke zkoumání hudby. Patří sem např. počítačová hudba, systematická muzikologie, MIR, *digital musicology* nebo *music informatics*.

## 1.1 Objektivizace analýzy interpretačního výkonu

Analýza interpretačního výkonu je mnohdy zkreslena subjektivním vnímáním posuzující osoby, např. hudebního kritika nebo muzikologa. Stejně tak je analýza ovlivněna zkušenostmi a znalostmi badatele. Bez exaktní analýzy je hudební tvar stále nezměřitelný a neprobadatelný. Pro přesné popsání interpretačního výkonu je potřeba analýza, která je schopna zachytit a popsat průběh interpretačního dění, ve vztahu k notovému zápisu a i nezávisle na něm, nelze se opírat pouze o popis slyšených interpretačních jevů.

Z českého muzikologického prostředí je významná práce od muzikologa Milana Kuny a elektroakustika Miloše Bláhy [16]. K výzkumu interpretačního výkonu použili prostředky tzv. kvantitativní analýzy, kdy pracují se zobrazením průběhu hudebního záznamu. Impulsem pro jejich bádání byla snaha přiblížit se k analýze hudebně

---

<sup>1</sup>U uvedených oborů v současné době neexistuje adekvátní terminologie, z toho důvodu je zachována anglická varianta.

interpretačního výkonu na základě objektivně měřitelných dat a exaktně prokazatelných výsledků. V rámci své práce zavedli specifickou českou terminologii týkající se interpretačního výkonu, např. délka trvání interpretačního výkonu, nepravidelný průběh hudby, míra rovnoměrnosti tempového pohybu nebo rozmezí estetické rezonance hudby. [16]

Britský muzikolog Nicolas Cook (\*1950) a americký hudebník a akademik José Antonio Bowen (\*1962) ve svých studiích zaměřených na analýzu interpretačního výkonu poukazují na potenciál využití moderních výpočetních metod<sup>2</sup>. Pokročilé postupy číslicového zpracování signálu umožňují zobrazit detailní a přesné informace o tempu a dynamice na rozdíl od práce s ručními stopkami nebo magnetofonovými pásky, což byl jediný způsob získávání dat v době analogových záznamů. Nicolas Cook se ve Velké Británii zasadil o formování oboru interpretační teorie (*performance studies*) a spolupodílel se na projektu při Royal Holloway University v Londýně s názvem *Research Centre for the History and Analysis of Recorded Music* (CHARM).

Význačnou prací výzkumného týmu CHARM je *Mazurka Project*<sup>3</sup>, která se zabývá vystižením rysů interpretace mazurek skladatele Frédérica Chopina (1810–1849). Cílem tohoto projektu bylo prozkoumat potenciál výpočetních metod pro extrakci hudebních dat a pro stylistickou analýzu interpretace z hudebních nahrávek. Chopinovy mazurky byly vybrány záměrně, jedná se o velmi oblíbené skladby, které jsou často interpretovány, a je dostupné velké množství jejich nahrávek. Také se jedná o útvar, který je interpretován charakteristickým způsobem, jenž však není v notovém zápisu zaznamenán. Ze závěrů se výzkumný tým snažil sledovat trend v historickém vývoji interpretace nebo charakterizovat různé typy přednesu na základě původu interpreta. Velká část práce se zaměřuje na extrakci informací o tempu a dynamice za pomoci metod MIR. Pro účely projektu byla vytvořena sada přídatných modulů<sup>4</sup> pro software Sonic Visualiser, který se stal jedním ze základních nástrojů pro výzkum (více v kapitole 4.1). [31]

Komplikací pro rozsáhlý výzkum může být velmi nízká dostupnost online databází nahrávek pro studijní nebo výzkumné účely. Problematické je získat např. nahrávky 1. poloviny 20. století. Často jsou dostupné jen kopie potřebných nahrávek a původní záznam je těžce dohledatelný.

Podrobněji budou metody objektivní analýzy interpretačního výkonu popsány v kapitole 2.1.

---

<sup>2</sup>Mezi významnou literaturu patří kniha Nicolase Cooka *Beyond the Score. Music as a Performance* (2013), které předcházela studie *Changing the musical object: Approaches to performance analysis* (2009). Studie José Antonia Bowena *Tempo, Duration and Flexibility: Techniques in the analysis of performance* (1996) je základní prací, kde autor využil postupy popisné statistiky.[6][4]

<sup>3</sup>*Mazurka Project*. Dostupné z URL:<<http://www.mazurka.org.uk/>>

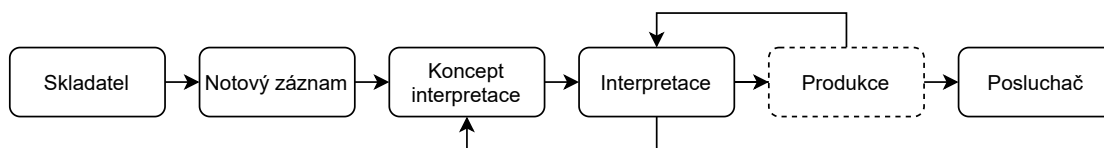
<sup>4</sup>*Mazurka Project Plugins for Sonic Visualiser*. Dostupné z URL: <http://sv.mazurka.org.uk/>

## 1.2 Předměty výzkumu interpretačního výkonu

Obecně se lze setkat s analýzou interpretačního projevu umělce v souvislosti s notovým zápisem hudebního textu, který je považován za jedinou „neměnnou“ veličinu mezi velkým množstvím interpretačních proměnných. [11]

Je třeba tedy rozlišovat mezi hudbou a jejím notovým zápisem. Notový zápis v tradici západní hudební kultury obsahuje informaci o výšce a relativní délce každé noty<sup>5</sup>, téměř vždy také obsahuje informace o dynamice. Dále se zde mohou vyskytovat instrukce k charakteru nebo specifickému stylu interpretace. U pozdějších až současných notových zápisů lze sledovat tendence k detailnějším a explicitnějším interpretačním instrukcím, pravděpodobně z toho důvodu, aby autoři odstranili nespecifikované úseky či nejednoznačnosti předpisu [18].

Na obr. 1.1 je uvedeno schéma tzv. hudební komunikace. Ukazuje dílčí složky procesu interpretace, jež sestává jak ze skladatele, který sdílí své hudební myšlenky, z interpreta, který tyto myšlenky ztvárňuje, tak i z posluchače, který je zase individuálně přijímá a hodnotí.



Obr. 1.1: Komunikační hudební řetězec, od skladatele k posluchači.

Avšak ani notový zápis nemusí být spolehlivou referencí vzhledem k různým vydáním podle odlišných edičních zásad<sup>6</sup>, a tak nelze interpretův projev posuzovat podle not, z kterých možná vůbec nehrál. K analýze lze přistupovat několika způsoby, nejjednodušším zkoumáním dodržování nebo nedodržování notového zápisu, nebo aplikací systému pravidel hudebního teoretika Heinricha Schenkera (1868–1935), která vycházejí ze struktury hudebního zápisu a ze zkušeností s realizací zapsaných hudebních jevů.

Každá hudební interpretace je vysoce individuální jak z produkčního hlediska, tak i z hlediska posluchače. Lze však uvést celou řadu parametrů či vlivů, které interpretaci zásadně ovlivňují: [18]

- **základní interpretační pravidla:** jedná se o konvence nebo normy, podle kterých se řídí většina interpretací, v opačném případě by interpretace byla považována za neobvyklou a zvláštní;

<sup>5</sup>V rámci práce jsou vynechány speciální případy notace jako jsou např. grafické partitury.

<sup>6</sup>Za ideální lze považovat odborná nebo kritická vydání, ve kterých bylo dílo zachováno tak, aby zůstalo co nejblíže původním záměrům skladatele.

- **interpretační plán:** performerův koncept interpretace, který může být ovlivněn následovně:
  - *interpretace hudební struktury a frázování*;
  - *moment překvapení* (odchylky od tradičního způsobu interpretace);
  - *stylistický a kulturní kontext a pravidla*;
  - *hudební nálada*, kterou se interpret rozhodne přenést na posluchače;
  - *místo provedení* (očekávané publikum, koncert / studiová nahrávka apod.);
- **interpretovo osobní, sociální a kulturní pozadí:** dosavadní zkušenosti, hudební škola a učitelé, přístup apod.;
- **fyzické vlivy:** fyzické schopnosti (hybnost prstů) a limity (přesnost, práce s dechem apod.);
- **zkoušení:** doba zkoušení hudebního díla (zkoušky mohou být také považovány za dílčí interpretace);
- **bezprostřední vlivy:**
  - *zpětná vazba* (zvuk nástroje, akustická odezva místnosti, odezva publika);
  - *externí vlivy* (vlhkost vzduchu, teplota okolí, rozptýlení);
  - „*vnitřní*“ *vlivy* (emoční rozložení interpreta, fyzická kondice).

Parametry interpretačního výkonu lze rozdělit do stejných základních kategorií, které jsou používány pro popis hudebního signálu obecně: čas (tempo, délka trvání), síla tónu (hlasitost, dynamika), barva tónu (témbr) a výška tónu (intonace, artikulace).

### 1.2.1 Tempo

Na úvod je vhodné definovat základní pojmy, které mohou být ve spojitosti s tempem zaměňovány.

Tempo označuje rychlost, kterou se skladba hraje, obecně představuje počet dob za minutu. Metrum označuje v hudbě střídání přízvučných a nepřízvučných dob. Je rozlišováno metrum dvoudobé, třídobé, čtyřdobé apod. Metrum v hudbě je vnímáno jako pulzace skladby a nemusí být vždy zřetelně zaznamenáno v notovém zápisu. Rytmus je střídání tónů různých délek v určitém vzorci a vždy je plně vypsán v notovém záznamu. Metrum a rytmus se vzájemně doplňují a v jejich souvislosti mluvíme o metrickorytmických vztazích. Metrum je také často uváděno jako synonymum pro taktové předznamenání, které určuje, kolik dob obsahuje jeden takt.

Tempové označení tedy v notách předepisuje, jak rychle se má skladba hrát, kde má zrychlit či zpomalit. Tempo a jeho změny se označují přibližně slovními výrazy, většinou italským názvoslovím, nebo přesně daným *metronomickým* údajem,



např. ♩ = 80. Číslo označuje počet úderů za minutu<sup>7</sup> a nota určuje, jaké hodnotě noty odpovídá časový úsek mezi dvěma údery, zda se jedná o notu půlovou, čtvrtovou, osminovou apod. Přehled běžných tempových označení je uveden v tab. 1.1. Specifickým označením je *ad libitum* (dle libosti), kdy je interpretovi ponechána absolutní volnost ve volbě tempa. [36]

Tab. 1.1: Základní tempová označení

BPM	italské názvosloví	český překlad
50	<i>lento</i>	rozvláčně, zdlouha
54	<i>adagio</i>	pomalů, zvolna
63	<i>andante</i>	volně krokem
88	<i>moderato</i>	mírně
132	<i>allegro</i>	vesele, rychle

### 1.2.2 Dynamika

Dynamika v hudbě označuje sílu (resp. hlasitost) přednesu. V zápisu klasické notace je dynamika zaznamenávána pomocí grafických značek neboli dynamických znamének, vedle toho se užívá také slovních výrazů a jejich zkratk. Základní přehled je uveden v tab. 1.2. Zajímavým jevem při interpretaci je interpretova tendence zrychlovat při zesilování a naopak zpomalovat při zeslabování. [36]

Ve většině případů je v notovém zápise dynamika jen hrubě naznačena nebo chybí úplně. Je přednostně na interpretovi, jak bude s hlasitostí pracovat na základě své volby frázování.

### 1.2.3 Témbr

Témbr (z francouzského *timbre*) neboli barva tónu je základní vlastností hudebního signálu. Označuje spektrální složení tónu, počet a intenzitu částkových tónů, šumů a formantů. Barva hudební nahrávky je utvářena několika faktory, prvotně instrumentací skladby samotné, vedle toho pak stavbou nástroje, jeho rozměry a materiálem nebo způsobem tvorby tónu. Interpret také ovlivňuje výslednou barvu svým jedinečným stylem, další významnou roli sehrává jeho věk, fyzické proporce (ovlivňující např. styl dýchání nebo motoriku), původ (respektive hudební škola, kde studoval). Dalšími okolnostmi jsou místo provedení, rozměry a rozvržení místnosti, její akustické vlastnosti, mezi které patří např. rozměry a rozvržení místnosti, její rezonance a módy, pohltivost použitých materiálů nebo atmosférické podmínky (vlhkost vzduchu).

<sup>7</sup>V praxi se setkáváme s označením MM (Mälzlův metronom) nebo BPM (Beats Per Minute – počet úderů za minutu).

Tab. 1.2: Základní dynamická označení

zkratka	italské názvosloví	český překlad
<b><i>ppp</i></b>	piano pianissimo	co nejslaběji
<b><i>pp</i></b>	pianissimo	velmi slabě
<b><i>p</i></b>	piano	slabě, tiše
<b><i>mp</i></b>	mezzopiano	středně slabě
<b><i>mf</i></b>	mezzoforte	středně silně
<b><i>f</i></b>	forte	silně
<b><i>ff</i></b>	fortissimo	velmi silně
<b><i>fff</i></b>	forte fortissimo	co nejsilněji
<b><i>fp</i></b>	forte piano	důraz na první notu
<b><i>cresc.</i> (&lt;)</b>	crescendo	zesilovat
<b><i>decresc.</i> (&gt;)</b>	decrescendo	zeslabovat
<b><i>sf, sfz</i></b>	sforzato	s důrazem

Při analýze hudebních nahrávek je základním činitelem měnícím barvu zvuku samotné pořízení záznamu. Zvuk ze záznamu nemusí vždy odpovídat tomu zvuku, který byl původně nahrán. Svou roli hraje rozmístění a volba mikrofonů, každý nástroj pak nebude snímán se stejnou intenzitou, dozvuk (přirozený nebo umělý) zkreslí frekvenční spektrum. Následný proces editace pak do nahrávky může vnést další zkreslující elementy (přidání komprese, ekvalizace apod.).

### 1.2.4 Výška tónu a artikulace

Obecně notový zápis definuje výšky tónu, které mají být v rámci skladby zahrány. Parametry interpretace založené na výšce tónu však zahrnují expresivní techniky, jako je např. *vibrato* nebo úmyslnou (či neúmyslnou) volbu intonace (např. míra „rozladěnosti“ při interpretaci klezmeru<sup>8</sup>).

Termín artikulace označuje, jakým způsobem po sobě následují tóny hudební skladby. Artikulace může být zapsána do notového zápisu, a to formou různých symbolů nebo pokynů, jak se má určitá pasáž ve skladbě provádět. Přehled základních pojmů je uveden v tab. 1.3

Při posuzování provedení hudebního díla je často věnována pozornost míře a způsobu užitého *vibrata*. *Vibrato* vzniká rychlým střídáním dvou tónů nepatrně rozdílných výšek, jedná se o frekvenční modulaci tónu. Vibrato je posuzováno především u zpěváků a hráčů na smyčcové nástroje. Např. mezi profesionálními a amatérskými zpěváky lze pozorovat signifikantní rozdíl v rozsahu intervalu vibrata [9]. Jiným jevem je *tremolo*, které označuje rychlé opakování téhož tónu, akordu nebo úderu na buben, a jedná se o amplitudovou modulaci tónu.

<sup>8</sup>Klezmer je označení pro hudební žánr světské židovské hudby.

Tab. 1.3: Přehled základních pojmů artikulace

italské názvosloví	český překlad	význam
legato	vázaně	tóny následují těsně za sebou bez mezer, značí se obloučkem spojujícím různé noty;
tenuto	oddělovaně	plné držení hodnoty noty s malým přízvukem a oddělením, značí se tečkou pod nebo nad notou;
staccato	krátce	tóny se hrají krátce a mezi nimi vznikají pomlky, značí se vodorovnou čárkou pod nebo nad notou;
glissando, gliss	sklouznutí	plynulý přechod z jednoho tónu na druhý, značí se čárkou nebo vlnovkou spojující dva tóny různé výšky.

Jak bylo zmíněno v kapitole 1.2.3, záznam nahrávky může zkreslovat výslednou barvu interpretace. Samotné pořízení nahrávky pak může mít vliv i na artikulaci a stylové aspekty interpretace. Při nahrávání ve studiu, nebo obecně při nahrávání bez diváků (tj. nejedná se o záznam živého vystoupení), se interpretace může v procesu nahrávání vyvíjet. Interpretovi je umožněno přehrát jeho přednes a následně může rozhodnout, co ve své hře upravit. Tím se může snažit vědomě přiblížit k charakteristickým prvkům jiného interpreta, jiné školy atd. Další vliv na nahrávku pak může mít přítomný producent. Tato skutečnost může komplikovat snahy klasifikovat nahrávky do skupin podle národnosti apod.

## 2 Music Information Retrieval

Music Information Retrieval (MIR), čili získávání informací o hudbě, v sobě spojuje předměty číslicového zpracování signálu (DSP), zkoumání muzikologie, strojového učení, získávání informací (*Information Retrieval*, IR) a statistiky. Obor je to relativně mladý, historie sahá do 90. let 20. století. První mezinárodní sympozium věnující se MIR (*International Symposium on Music Information Retrieval*, ISMIR) proběhlo v říjnu 2000, a nyní se jedná o každoroční událost s více jak stovkou prezentovaných příspěvků. V současné době je to rychle se rozvíjející oblast výzkumu, především díky potřebám aplikací streamování hudby, jako jsou Spotify, iTunes nebo Pandora<sup>1</sup>. Rozvoj je samozřejmě podporován rychle se zdokonalující technikou a zvyšujícím se výkonem osobních počítačů.

MIR spadá do stejné oblasti výzkumu jako zpracování řeči. Mnoho technik je původně vytvořeno pro analýzu řeči a později aplikováno na hudební signály. Ty se však od řeči podstatně liší a mají specifické akustické a strukturní znaky, které je odlišují od mluveného slova a nehudbních signálů.

Výzkum v oblasti MIR je soustředěn zejména na extrakci vhodně zvolených parametrů z hudby. Parametry jsou extrahovány zpravidla z hudebních nahrávek, ale také ze symbolických reprezentací, jako je notová transkripce nebo MIDI zápis. Díky získaným parametrům pak lze vytvářet systémy, které jsou schopny třídit nahrávky dle žánrů nebo nálady. Typickými úlohami MIR je doporučování hudebního obsahu, transkripce hudebního signálu do notového zápisu, synchronizace nahrávky s notovým záznamem, rozpoznávání hudebního nástroje, určení hudební tóniny nebo hudební struktury atd. Extrahované parametry rozdělujeme do tří úrovní:

- nízkoúrovňové parametry (*low-level features*),
- středněúrovňové parametry (*mid-level features*),
- vysokoúrovňové parametry (*high-level features*).

Systémy vyhodnocující nízkoúrovňové parametry posuzují elementární bloky hudební nahrávky. Tyto parametry lze získat přímým výpočtem z hudebního signálu a popisují základní akustické vlastnosti, jako je frekvence, délka trvání tónů, spektrum tónu nebo intenzita tónu. Středněúrovňové reprezentace charakterizují hudební strukturu signálu, jako je rytmus nebo harmonie. Patří sem např. pokročilejší metody odhadu tempa (*tempo estimation*), fluktuace tempa nebo onset detekce. Vysokoúrovňové parametry poskytují nejkomplexnější informace o hudební nahrávce, jako je rozpoznání hudebního nástroje, určení hudebního žánru nebo emoce.

---

<sup>1</sup>Streamovací služba Pandora je v současnosti dostupná pouze v USA.

## 2.1 Music Performance Analysis

Music Performance Analysis (MPA, analýza interpretačního výkonu) je pole výzkumu ležící na hranici MIR, muzikologie a psychoakustiky. Obecně patří k okrajovým tématům, kterými se MIR komunita zabývá, v současnosti se jí však dostává větší a větší pozornosti. Oblast MPA se zaměřuje na komplexní studium interpretačního výkonu, zabývá se pozorováním, extrakcí, popisem, vyložáním a modelováním hudebních a MIR parametrů.

Velká část prací se soustředí na explorativní metody pro analýzu hudebních nahrávek a popis interpretačních charakteristik. Tyto studie typicky extrahují různé parametry pro nalezení společných rysů nebo odlišností mezi nahrávkami. Mezi výstupy patří charakteristiky, jako je křivka průběhu tempa, hlasitostní křivka nebo histogram. Některé práce usilují o nalezení základních vzorců mezi interpretačními parametry, které by mohly být užitečné při identifikaci trendů měnících se skrze různá období [10][25].

Hlavním bodem v systematickém výzkumu interpretačního výkonu je proces získávání dat pro následnou analýzu. Mezi různými metodami, které byly navrženy a použity pro získání dat, se objevují dva základní přístupy: monitorování interpretace (či rovnou interpretačních parametrů) technickými přístroji, nebo extrakce parametrů z hudebních nahrávek interpretace. Oba přístupy mají své výhody a nevýhody.

Metoda monitorování nese výhody přesných a detailních výsledků, jelikož měřicí zařízení mohou sledovat parametry víceméně přímo. Nevýhodou však je, že taková analýza se omezuje pouze na provedení uskutečněná za speciálních podmínek a pouze s vybranými interprety. [26]

Extrakce parametrů z hudební nahrávky může směřovat k méně přesným datům. Mimo to, některé parametry je z nahrávky téměř nebo úplně nemožné získat, např. informace o užívání klavírového pedálu nebo pohyb interpreta (sledovat pohyb může být žádoucí např. u smyčcového kvarteta, kdy je tento parametr slučitelný s provázaností souboru jako celku). Výhodou této metody je však aplikace na jakoukoliv dostupnou nahrávku. [18]

Většina dosavadních prací se věnuje převážně monofonním skladbám, z velké části pro sólový klavír. Jedním z důvodů je fakt, že klavír je velmi rozšířeným hudebním nástrojem s obrovským repertoárem sólových skladeb. Důležitější je však charakteristika jeho zvuku. Tóny produkované klavírem jsou perkusivního charakteru, což z klavíru činí nástroj vhodnější k přesnější analýze např. onset detekce než jiné nástroje, zejména smyčcové. Mimoto klavírní interpreti jsou v jistém ohledu více omezení než jiní hudebníci. Jsou limitováni klávesami, které rozehrávají struny naladěné na přesné frekvence (a v temperovaném ladění), což neumožňuje artikulaci,

jako je vibrato nebo chtěné rozladění tónu. Nemají také příliš velký vliv na výslednou barvu, tu mohou ovlivnit nejvýše způsobem úhozu, avšak po něm nemají nad tónem už žádnou kontrolu, kromě řízení délky tónu. Zmíněné skutečnosti vedou k předpokladu přesných výsledků extrakce parametrů a jejich následné analýzy. [18]

Příkladem aplikace výzkumu MPA je vývoj softwaru pro výuku hudby. Záměrem těchto nástrojů je doplnit fyzickou výuku o interaktivní zpětnou vazbu studentů analyzováním nahrávek z praktických lekcí. Jedním z cílů je, aby by byl „interaktivní hudební učitel“ schopen poukázat na problematické části studentova přednesu, poskytnout srozumitelnou analýzu a individualizovat studentův proces učení na základě chyb a obecného progresu. V současnosti je dostupných několik komerčních aplikací, které však nejsou plnohodnotnými nástroji, jež by dokázaly nahradit klasické vzdělávací prostředí. Mezi takové vyučující aplikace patří SmartMusic<sup>2</sup>, Yousician<sup>3</sup>, Music Prodigy<sup>4</sup> nebo SingStar<sup>5</sup>. [12][19]

Dalším neméně zajímavým využitím je oblast generativního modelování. Cílem je vygenerovat interpretaci (předpovědět průběh tempa a dynamiky, nebo přímo vygenerovat audio nahrávku) na základě poskytnutého notového materiálu jako vstupu. Příkladem takové práce je studie [21], která se za pomoci rekurentní neuronové sítě pokouší naučit generovat expresivní hudební přednes z notové reprezentace na základě předložených živých hudebních interpretací. Výsledkem je model, který je schopen vygenerovat škálu několika provedení stejného hudebního díla s interpretačními rozdíly. Dalším příkladem je práce [15], kde autoři aplikovali LSTM (*Long Short-Term Memory – dlouho-krátkodobá paměť*) rekurentní neuronovou síť pro generování hudebního výrazu pro MIDI soubory.

---

<sup>2</sup>SmartMusic. Dostupné z URL:<<https://www.smartmusic.com/>>

<sup>3</sup>Yousician. Dostupné z URL:<<https://yousician.com/>>

<sup>4</sup>Music Prodigy. Dostupné z URL:<<http://www.musicprodigy.com/>>

<sup>5</sup>SingStar je karaoke hra pro konzole PS4. Cílem je trefovat správné tóny zvolené skladby.

## 3 Parametrizace hudebního signálu

MIR parametry objektivně popisují vlastnosti hudebního signálu. Většinou se jedná o vektorové reprezentace, jejichž prvky odpovídají časové posloupnosti analyzovaného signálu, nebo skalární reprezentace, jejichž hodnota je vztažena k jednomu segmentu analyzovaného signálu.

Z časového hlediska lze parametry dělit na okamžité, úsekové nebo globální. Z hudebního hlediska je možné parametry rozdělit do skupin popisující tempo, dynamiku, tímbr a výšku tónu.

### 3.1 Parametry popisující tempo

Při srovnávání dvou interpretací stejné skladby je jedním z nejpozorovatelnějších rozdílů odlišně volené tempo, které tak patří k nejdůležitějším interpretačním parametrům. Pro účely analýzy interpretačního výkonu se nemusí jevit porovnávání rytmické nebo metrické struktury jako příliš vypovídající, jelikož ta u identických skladeb zůstává víceméně stejná. Lze však uvažovat, že by srovnávání metra a rytmu mohlo pomoci např. při analýze důrazů na různých dobách.

#### Odhad tempa

Tempo vyjadřuje rychlost, jakou je daná skladba přehrávána. Obecně lze tempo  $T$  určit jako poměr počtu úderů za minutu a délky časového úseku

$$T = \frac{B \cdot 60}{\Delta t} \quad [\text{BPM}], \quad (3.1)$$

kde  $B$  značí počet úderů a  $\Delta t$  vybraný časový úsek. V rámci interpretace skladby tempo však nezůstává konstantní, ale působením mnoha faktorů, jako je styl hry nebo interpretační zkušenosti, je značně proměnlivé.

V oblasti MIR je užíván pojem *globální tempo*, které určuje průměrné tempo hudební nahrávky. Obvykle je extrahováno u nahrávek, u kterých je předpokládán relativně stabilní průběh tempa a jeho hodnota je získána průměrováním jednotlivých hodnot tempa v čase.

Ne vždy však je k nahrávce dostupná informace o počtu úderů. Pro účely identifikace pozic dob slouží buď manuální vytvoření anotace nebo automatické metody. Jednou z možností automatických metod je aplikace **onset detekce**. Účelem onset detekce (*onset detection*) je rozpoznat fyzické začátky not nebo jiných hudebních událostí v hudební nahrávce. Podstatou je zachytit náhlé změny v hudebním signálu, které jsou typicky způsobeny počátkem nové události. Výsledkem je detekční

křivka (*novelty curve*), jejíž vrcholy indikují signifikantní změny signálu a tím pravděpodobnou přítomnost počátku nové události. Extrahované začátky not lze využít při detekci tempa, lokaci jednotlivých dob (*beat tracking*), automatické transkripci či segmentaci signálu.

V současnosti se v procesu onset detekce a beat trackingu aplikují čím dál častěji metody hlubokého učení. Např. v případě balíčku pro Python `madmom` jsou pro algoritmy beat trackingu využity mimo jiné modely rekurentní neuronové sítě nebo skrytý Markovův model.

## 3.2 Parametry popisující dynamiku

Je třeba rozlišit termíny intenzita a hlasitost neboli hladinu intenzity. Intenzita zvuku je fyzikálně měřitelná veličina a je definována podílem výkonu  $P$  zvukového vlnění a plochy  $S$ , kterou vlnění prochází. Hlasitost vyjadřuje subjektivní lidské vnímání zvuku. Fyziologie lidského ucha způsobuje, že člověk zvuk s lineárně rostoucí intenzitou vnímá logaritmicky. V logaritmické stupnici lze vyjádřit hladinu logaritmickým poměrem [28]

$$L_I = 10 \log \frac{I}{I_0} \quad [\text{dB}], \quad (3.2)$$

kde  $I$  je intenzita zvuku a  $I_0$  intenzita prahu slyšení.

### Efektivní hodnota signálu

Efektivní hodnota (*Root-Mean Square*, RMS) je základním parametrem popisujícím globální energii a dynamický vývoj hudebního signálu. Vyjádřena je vztahem [5]

$$\text{RMS}_t = \sqrt{\frac{1}{n} \sum_{i=1}^n s(k)^2}, \quad (3.3)$$

kde  $t$  je časový rámeček,  $s(k)$  je hodnota  $k$ -tého vzorku a  $n$  počet vzorků v časovém rámečku  $t$ .

### Křivka nízké energie

Křivka popisující nízkou energii (*low energy*) je využívána k odhadu rozložení energie v čase. Zobrazuje, zdali signál v dynamice zůstává víceméně konstantní, nebo jestli některé části vykazují větší kontrast. Obvykle je vyjádřena procentuálně jako poměr časových rámečků  $t$ , jejichž RMS energie klesla pod průměr vůči zbylým rámečkům [34]. Nahrávka vokální hudby s četnými pauzami tak bude mít vysokou hodnotu nízké energie, zatímco plynulé smyčcové pasáže budou mít hodnotu nízkou. Parametr je



vhodný např. při rozpoznávání hudebního žánru, při posuzování agogiky nebude mít pravděpodobně velkou vypovídající hodnotu.

### 3.3 Parametry popisující témbor

Z polyfonní hudby je lidské ucho schopno rozeznat jednotlivé nástroje právě díky rozdílné barvě zvuku. Tak jako je člověk schopen rozeznat celkovou barvu hudebního signálu, ale je schopen věnovat pozornost i dílčím dokumentům, tak i systémy MIR jsou schopny díky globální barvě zvuku rozeznat v nahrávce hudební žánr nebo náladu, nebo hudební signál detailně analyzovat a rozeznat např. jednotlivé nástroje.

Pro účely charakteristiky vývoje zvuku v čase je zaveden pojem ADSR obálka (attack, decay, sustain, release). Pro většinu akustických tónů je však těžko zachytitelná ideální obálka, často kvůli absenci fáze poklesu nebo držení tónu. Můžeme se tak setkat se zjednodušenou reprezentací, kde chybí fáze poklesu a fáze držení a uvolnění je sloučena. Barva zvuku je nejvíce ovlivněna fází náběhu, konkrétně dobou náběhu (*attack time*), sklonem náběhu (*attack slope*) a intenzitou náběhu (*attack intensity*). [27]

#### Počet průchodů nulovou úrovní

Parametr počet průchodů nulovou úrovní (*zero crossing rate*) udává, kolikrát protne signál osu  $x$  (tzn. nulovou úroveň) neboli kolikrát se změní znaménko hodnoty amplitudy v daném časovém rámci  $t$ .

Primárně se parametr používá v oblasti zpracování řeči nebo k rozpoznávání perkusivních zvuků, lze jej však zařadit i mezi parametry popisující barvu zvuku.

#### Spektrální centroid

Spektrální centroid (*Spectral Centroid*) určuje, kde se nachází střed modulu spektra. Popisuje tedy, kde leží těžiště spektrální energie. Obecně je průměrnou hodnotou spektra [24]

$$\mu = \frac{\sum_k S(k)f(k)}{\sum_k S(k)}, \quad (3.4)$$

kde  $S(k)$  je modul spektrálního koeficientu  $k$  a  $f(k)$  jeho střední kmitočet.

V rámci popisu barvy hudebního zvuku je spektrální centroid spojován s jasnou barvou zvuku neboli jasností zvuku (někdy označováno také jako ostrost zvuku) [18]. Nízké hodnoty spektrálního centroidu odpovídají menší jasnosti.

## Spektrální rozpětí

V návaznosti na spektrální centroid je zaveden parametr spektrálního rozpětí (*Spectral Spread*), který popisuje rozložení spektra kolem jeho průměrné hodnoty. Někdy je též označován jako okamžitá šířka pásma. Jedná se o druhý centrální moment, který ze statistického hlediska označuje hodnotu rozptylu  $\sigma^2$  [17]

$$\sigma^2 = \int (x - \mu)^2 f(x) dx . \quad (3.5)$$

Pokud parametr spektrálního rozpětí vykazuje nízké hodnoty, rozložení spektra je koncentrováno kolem spektrálního centroidu.

## Ujíždění spektra

Ujíždění spektra (*Spectral Roll-off*) určuje hraniční frekvenci, pod kterou se nachází předem určené procento energie signálu, většinou se jedná o 85 % nebo 95 %. Jedná se o jeden ze způsobů, jak odhadnout míru vysokých frekvencí obsažených v signálu. Nízké hodnoty ukazují na malou šířku pásma hudebního signálu. [17]

## Spektrální tok

Spektrální tok (*Spectral Flux*) nebo také spektrální proměnlivost je ukazatelem míry proměnlivosti výkonového spektra signálu. Parametr spektrálního toku lze např. využít při detekci řečového signálu nebo onset detekci v hudební nahrávce [24]. Lze jej spojit s vnímáním drsnosti zvuku [18].

## Melovské keprstrální koeficienty

Melovské keprstrální koeficienty (*Mel-Frequency Cepstral coefficients*, zkráceně MFCCs) popisují krátkodobé spektrum signálu. Původně byly využívány hlavně v oblasti zpracování řeči, řada studií však ukazuje jejich úspěšnost např. v rozpoznávání hudebních nástrojů nebo rozpoznávání hudebních žánrů. Obecně MFCCs parametry charakterizují tvar spektrální obálky audio signálu [18].

Kepstrum je definováno jako zpětná Fourierova transformace (případně diskrétní kosinová transformace) z logaritmovaného modulu spektra signálu [3]. Mel kepstrum je kepstrum, u kterého bylo nejprve modulové spektrum signálu filtrováno bankou melovských filtrů. MFCC koeficienty jsou pak koeficienty mel kepstra [27].

Jednotka mel (z anglického *melody*) označuje relativní výšku tónu. Rozlišujeme absolutní a relativní výšku tónu, většina lidí je však schopna rozeznat pouze relativní výšku tónu. Ke každému tónu byl experimenty určen tón, jehož výška se zdá poloviční, dvojnásobnou atd. V takto sestavené stupnici byla jednotka pro subjektivní výšku tónu nazvána mel. Závislost mezi reálnou frekvencí a subjektivní frekvencí se

nazývá melová stupnice (*mel scale*). Průběh závislosti je přibližně lineární do frekvence 1 kHz a logaritmický nad frekvencí 1 kHz. Člověk je např. schopen pohodlně určit rozdíl mezi zvukem o  $f = 100$  Hz a zvukem o  $f = 200$  Hz. U vyšších frekvencí už tomu tak ale není. Pro člověka je mnohem těžší rozlišovat mezi vyššími frekvencemi, takže i když např. mezi frekvencemi  $f = 1000$  Hz a  $f = 1100$  Hz je stejná vzdálenost, jako v prvním případě, lidské ucho tuto vzdálenost stejně nevnímá. Skutečnost, že MFCC parametry simulují vnímání zvuku člověkem je důvodem, proč jsou základními parametry např. pro strojové učení a klasifikaci.

Subjektivní výška tónu také závisí na jeho hlasitosti. Při zvyšování hlasitosti tónů s nízkým kmitočtem se jejich subjektivní výška zmenšuje, u tónů s frekvencí vyšší než 4000 Hz se ale zvyšováním hlasitosti jejich subjektivní výška nezměňuje [28].

Vstupní signál je nejprve za pomoci Fourierovy transformace převeden do frekvenční oblasti. Poté je získáno modulové spektrum signálu, které je následně filtrováno bankou trojúhelníkových filtrů s nelineárním rozložením frekvencí, tzv. bankou melovských filtrů. Tyto filtry jsou rozmístěny na melové stupnici (*Mel-frequency scale*), která přiřazuje subjektivně vnímanou frekvenci tónu k jeho reálné změřené frekvenci. Jedním ze způsobů, jak sestavit melovou stupnici, je

$$\text{Mel}(f) = 2595 \log \left( 1 + \frac{f}{700} \right), \quad (3.6)$$

kde  $f$  je frekvence vyjádřena lineárně a  $\text{Mel}(f)$  je frekvence vyjádřená v melové stupnici.

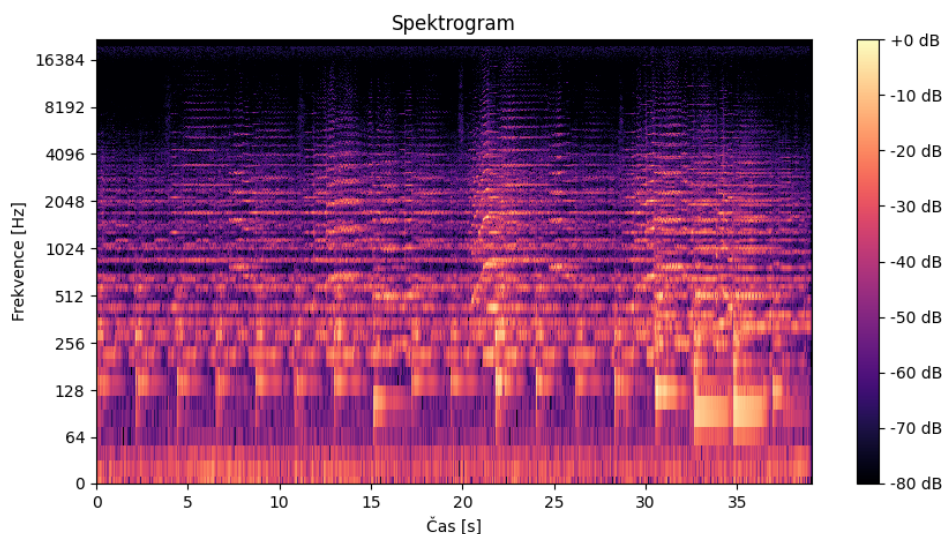
Nakonec je pro každý filtr vypočítán logaritmus energie a výsledný vektor logaritmických hodnot je za pomoci diskrétní kosinové transformace převeden zpět do časové oblasti. Výsledkem je vektor melovských koeficientů, které kompenzují nelineární vnímání frekvencí lidským uchem.

## Spektrogram

Spektrogram je vizuální reprezentací kmitočtového spektra v čase. Jedná se o vyobrazení modulu krátkodobé Fourierovy transformace signálu (STFT), jejímž principem je segmentace signálu do kratších úseků, kdy je provedena rychlá Fourierova transformace pro každý segment zvlášť [29].

Na výsledné zobrazení má zásadní vliv velikost zvoleného váhovacího okna k segmentaci a jeho zvolená délka. Vzhledem k logaritmickému vnímání zvuku lidským uchem je frekvenční osa vynesena logaritmicky. Příklad spektrogramu je uveden na obr. 3.1.

Spektrogram lze primárně využít pro kontrolu intonace s pomocí notové partitury. Zobrazení jednotlivých frekvencí zastoupených ve spektru pak umožňuje lépe



Obr. 3.1: Spektrogram nahrávky realizované souborem Smetanova kvarteta (1959).

určit pozice tónů i v polyfonických nahrávkách. Ve spektrogramu lze také pozorovat jevy, jako je např. *tremolo* nebo *vibrato* (viz kapitola 1.2.4). Tyto úkazy jsou běžné pro smyčcové nástroje nebo pro zpěv. Obecně se jedná o lokální změny v intenzitě a kmitočtu, nemusí nezbytně vyvolávat pocit změny vnímané hlasitosti nebo výšky tónu. Tóny s vibratem jsou pak obecně sluchem vnímány jako zvukově bohatší a plnější. Takto modulované tóny totiž obsahují ve spektru nové složky.

Obměnou spektrogramu je mel spektrogram. Jedná se o spektrogram, kde jsou hodnoty frekvencí na ose  $y$  převedeny do melové stupnice (více v kapitole 3.3).

### 3.4 Parametry popisující výšku tónu

Detekce melodie v hudební nahrávce (tzv. *pitch tracking* nebo *f0 estimation*) je stále velkou výzvou v oblasti výzkumu MIR. Zatímco určení vývoje melodie v nahrávkách jednoho hudebního nástroje nebo hlasu je již možné vcelku přesně, tak v polyfonických strukturách je přesné určení výšky tónu komplikováno přítomností několika paralelních hlasů.

Rozšířenou metodou pro určení výšky tónu je krátkodobá Fourierova transformace (STFT), respektive její vizualizace ve formě spektrogramu, který byl blíže popsán v kapitole 3.3 [24]. Další možností je konstantní  $Q$  transformace (*Constant Q-transform*, CQT), která může vykazovat lepší rozlišovací schopnosti než STFT. [22]

Problémem při detekci melodie může být rozdílné naladění nástrojů skrz různé interpretace. Hudební nástroje se při ladění řídí tzv. *komorním a*. Jedná se o me-

zinárodně platný referenční tón  $A_4$  s frekvencí 440 Hz. Ovšem přesná frekvence, na kterou jsou nástroje naladěny, se může měnit v závislosti na tom, zda se jedná o historické hudební nástroje (typicky barokní nástroje se ladí na 415 Hz), či na preferenci barvy zvuku. Detekce ladicí frekvence může pomoci zajistit spolehlivější výsledky při detekci melodie.

Odchylka výšky tónu (*pitch deviation*) sleduje odchýlení výšky tónu v hudební nahrávce od referenční výšky tónu, tzn. sleduje správnost intonace podle notového zápisu. Za referenční kmitočet lze považovat údaj o výšce tónu v notovém zápise. Na prvním místě je odhadnuta základní frekvence hudebního signálu. Odchylka výšky tónu je pak určena jako rozdíl odhadnuté a referenční frekvence [26]

$$\delta f = f_i - f_{\text{ref}} \quad [\text{Hz}] , \quad (3.7)$$

kde  $f_i$  je změřená frekvence a  $f_{\text{ref}}$  jí odpovídající referenční frekvence v notovém zápise.

### Chroma parametry

*Chroma* parametry odkazují na rozdělení tónů do dvanácti tříd, které jsou odvozeny od rozložení dvanácti tónů ve stupnici podle temperovaného ladění. Jednotlivé tónové třídy jsou znázorněny v tabulce 3.1.

Tab. 3.1: Tónové třídy

1	2	3	4	5	6	7	8	9	10	11	12
C	C $\sharp$	D	D $\sharp$	E	F	F $\sharp$	G	G $\sharp$	A	A $\sharp$	H

Chroma parametr je reprezentován vektorem s dvanácti prvky pro daný časový rámec  $t$ . Chromagram je pak vizualizací tónového obsahu nahrávky v čase. Ukázka chromagramu je uvedena na obr. 6.2. Chroma parametry jsou vhodné k analyzování hudby, která vznikla v prostředí západního harmonického myšlení, tedy její notový obsah lze kategorizovat právě do dvanácti tónových tříd. Hlavní výhodou chroma parametrů je schopnost zachytit harmonickou a melodickou strukturu hudby i při proměnlivé instrumentaci a barvě zvuku. Typickým příkladem využití chroma parametrů je provedení synchronizace identických nahrávek z databáze pomocí algoritmu borcení časové osy (DTW) nebo detekce akordů a harmonické struktury hudebního signálu.

## 4 Softwarové nástroje pro analýzu interpretačního výkonu

Nejmodernější MIR algoritmy pro zpracování hudebního signálu sestávají většinou ze dvou částí. Nejprve jsou z nahrávky extrahovány nízkourovňové parametry ( fáze extrakce parametrů) a následně jsou získané informace analyzovány ( fáze analýzy parametrů). Mezi nejsoučasnější přístupy patří využití metod strojového učení a neuronových sítí ve fázi analýzy parametrů.

Pro výzkum a rozvoj aplikací byla navržena řada softwarových knihoven. Některé se soustředí převážně jen na extrakci nízkourovňových parametrů z hudebních signálů, např. Marsyas, YAAFE nebo openSMILE. Další knihovny zahrnují i některé středněúrovňové parametry (např. onset detekce nebo beat tracking). Mezi takové patří např. sada nástrojů pro MATLAB `MIRtoolbox`, C++ knihovna `Essentia` nebo Python knihovna `librosa`. Metody strojového učení využívá např. Python knihovna `madmom`.

### 4.1 Sonic Visualiser

Sonic Visualiser je velmi sofistikovaná a uživatelsky přívětivá aplikace pro analýzu, vizualizaci a anotaci hudebních nahrávek. Vzhledem a rozvržením uživatelského prostředí připomíná softwaru pro editaci hudby. Velkou výhodou je podpora pluginů pro rozšíření funkcí, významnými jsou pluginy formátu Vamp<sup>1</sup>, které umožňují řadu automatických metod pro hudební analýzu.

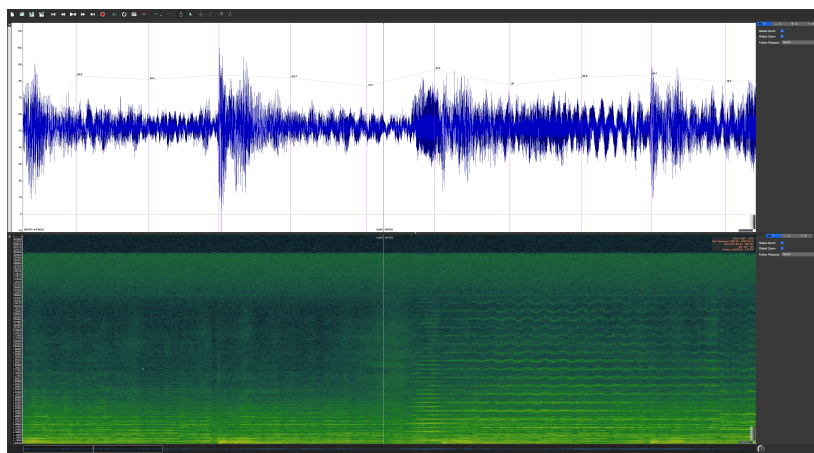
Aplikace vznikla na škole Queen Mary University v Londýně. Je psaná v jazyce C++. Cílem bylo vytvořit nástroj pro jednoduchou vizualizaci hudebních informací a umožnit detailní hudební analýzu badatelům, kteří se nepohybují v programátorském prostředí. Mezi základní vizualizační nástroje patří grafické zobrazení časového průběhu zvukových dat a spektrogramu, jehož výchozí parametry jsou nastaveny vhodně pro hudební analýzu. Aplikace také zahrnuje jednoduché metody pro anotaci hudebních souborů (např. označování pozice dob za pomoci myši, klávesnice nebo externího MIDI zařízení). Sonic Visualiser používá k zobrazování systém panelů a vrstev, který uživateli umožňuje přehledně srovnávat výsledky různých analytických metod či různé nahrávky. Ukázka zobrazení panelů je na obr. 4.1. Každý panel má svůj vlastní souřadnicový systém a každá vrstva může obsahovat rozdílné formátovací parametry, zobrazovat odlišná data získaná analýzou zvukového souboru

---

<sup>1</sup>Vamp je formát zásuvných modulů pro zpracování hudebních signálů. Na rozdíl např. od VST pluginů na výstupu není hudební signál, ale parametry, typicky časové hodnoty, spektrogramy nebo průběh amplitudové obálky. Více zde: <<https://vamp-plugins.org/>>

nebo může obsahovat komentář dat či další informace určené k zlepšení orientace v souboru. [7]

K nástroji Sonic Visualiser se vážou další tři aplikace: Sonic LineUp, Tony a Sonic Annotator. Sonic LineUp je nástroj pro rychlou synchronizaci a následnou vizualizaci rozdílných hudebních nahrávek stejného hudebního obsahu. Aplikace Tony je navržena pro analýzu melodie v monofonních nahrávkách. V případě Sonic Annotator se jedná podobný nástroj jako Sonic Visualiser avšak bez interaktivního prostředí v podobě konzolové aplikace, což umožňuje využití tohoto v jiných projektech. Je vhodný pro dávkové zpracování hudebních nahrávek a získání jejich hudebních parametrů. Zároveň podporuje pluginy formátu Vamp, stejně jako Sonic Visualiser. [7]



Obr. 4.1: Uživatelské prostředí Sonic Visualiser.

## 4.2 MIRtoolbox

K význačným projektům patří projekt z finské University of Jyväskylä, který nese název MIRtoolbox. Jedná se o sadu nástrojů pro programové prostředí MATLAB. Nabízí celou řadu funkcí umožňující extrakci hudebních parametrů z hudebních nahrávek, od běžných nízkourovňových až po vybrané vysokoúrovňové.

## 4.3 AMPACT

Dalším nástrojem pro analýzu interpretačních dat je AMPACT<sup>2</sup> (The Automated Music Performance Analysis and Comparison Toolkit). Jedná se rovněž o sadu nástrojů pro programové prostředí MATLAB. AMPACT umožňuje automatickou analýzu nahrávek, ke kterým je dostupný notový zápis. Využitím informace z notového záznamu je

---

<sup>2</sup>AMPACT. Dostupné z URL:<<https://ampact.tumblr.com/>>

system schopen předpokládat sled hudebních událostí. Následující verze by měla být schopna poskytnout takové nástroje pro extrakci dat z polyfonních nahrávek, aby bylo možné porovnávat různé interpretace stejného hudebního díla.

## 4.4 librosa

Řada nástrojů a skriptů pro MIR vzniká v prostředí MATLAB (MIRtoolbox) nebo C++ (např. openSMILE od společnosti audeERING), ale vedle toho vzrůstá zájem o využívání jazyka Python. Výhodou je řada dostupných kvalitních knihoven pro strojové učení (např. `scikit-learn`<sup>3</sup>), nebo rozsáhlý katalog balíčků pro práci s textovými daty.

To byl jeden z impulsů pro vytvoření balíčku `librosa`, knihovny pro zpracování audio signálů, která byla vyvinuta na americké Columbia University. Obsahuje základní stavební bloky nezbytné pro tvorbu MIR systémů. Knihovna `librosa` sestává z několika submodulů, zde jsou uvedeny některé z nich [20]:

- `librosa.core` – sada základních funkcí pro hudební analýzu a pro manipulaci se soubory (např. načtení souboru z disku);
- `librosa.display` – sada nástrojů k různému typu zobrazování v časové nebo frekvenční oblasti;
- `librosa.feature` – sada funkcí pro extrakci nízkoúrovňových parametrů;
- `librosa.onset` – sada funkcí pro výpočet onset detekce;
- `librosa.beat` – sada funkcí pro beat-tracking;
- `librosa.decompose` – obsahuje funkce pro oddělení jednotlivých hlasů v polyfonické nahrávce;
- `librosa.effects` – sada nástrojů pro zvukové efekty.

## 4.5 madmom

Stejně jako `librosa` je knihovna `madmom` napsána v jazyce Python. Její největší předností je využívání metod strojového učení za účelem získání komplexních informací z hudební nahrávky, např. počátků hudebních událostí, tempa, pozic dob, klavírní transkripce aj. Knihovna `madmom` je vystavena na třech rozšířených knihovnách: NumPy, SciPy a Cython. [2]

---

<sup>3</sup>open source knihovna s jednoduchými a efektivními nástroji pro prediktivní analýzu dat, vystavěná na knihovnách NumPy, SciPy a matplotlib



## 5 Databáze

Pro účely diplomové práce byla vytvořena databáze sestávající z 31 nahrávek *Smyčcového kvartetu č.12 F dur* (1893) od českého romantického skladatele Antonína Dvořáka (1841–1904) odlišné datace a od interpretů různého původu. Většina ansámblů jsou mezinárodně uznávanými hudebními tělesy, některé se svým repertoárem zaměřují přímo na dvořákovskou tvorbu. Přehled všech nahrávek je uveden v příloze B. Jednotlivé nahrávky jsou předzpracovány následujícím způsobem: všechny jsou ve formátu .wav (Waveform Audio File Format), mají sjednocenou vzorkovací frekvenci  $f_{vz} = 44,1$  kHz a bitová hloubka je rovna 16 bitům. Kromě toho byla provedena konverze ze stereo formátu do mono.

### 5.1 Výběr motivu

Analýze byly podrobeny nahrávky jak globálně, tak lokálně. Podrobněji byl analyzován motiv ze 2. věty *Lento*. Jedná se o takt 1–10, kde 1. housle zaujímají vedoucí pozici a ostatní nástroje mají doprovodnou funkci. Notová transkripce je k dispozici v příloze C. Předlohou je verze dostupná ke stažení z volně přístupné online notové databáze IMSLP: Petrucci Music Library [35].

Metrum je 6/8, tempové označení je zapsáno jako *lento* (rozvláčně, zdlouha, viz tab. 1.1). Melodická linie 1. houslí má předpis *molto espressivo* (velmi expresivně), lze tedy předpokládat, že dynamika i tempo bude mít četné výchyly od základního předpisu. V úvodu nástroje hrají v dynamice *piano*, na konci taktů se pak objevuje pro všechny nástroje *crescendo/decrescendo*, maximální předepsanou dynamikou je *mezzoforte*.

V dostupných nahrávkách se věta *Lento* vyznačuje velmi odlišnou expresivitou a rozdílně voleným tempem. Některé ansámble volí spíše stabilnější průběh skladby, s výsledným dojmem hry „podle metronomu“, zatímco některé nahrávky jsou charakteristické velmi kolísavým tempem navázaným na kontrastní dynamiku a expresivní projev.

### 5.2 Anotace

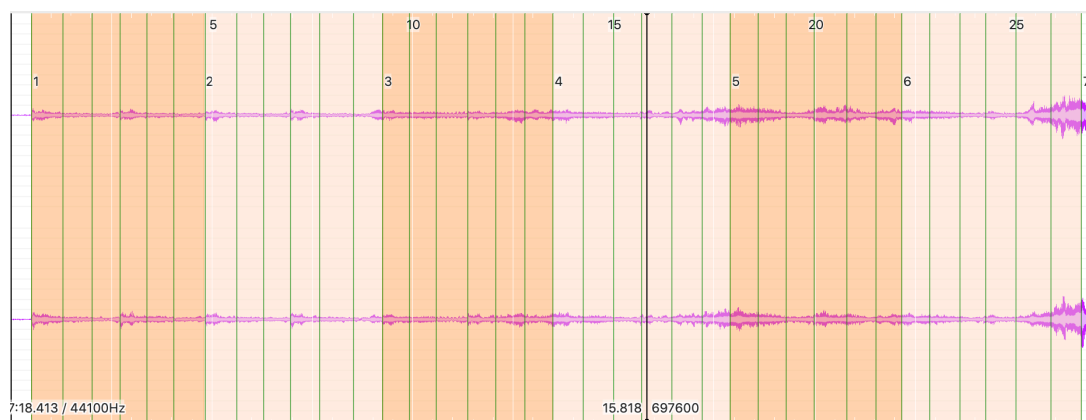
Pro označení notové události v nahrávce lze využít několika metod. Jednou možností je během nahrávky stisknout (angl. *tap*) určenou klávesu na klávesnici (nebo jiném vstupním zařízení) v daný čas a tím dát systému informaci, kde se daná notová událost nachází na časové ose. Tento proces se obecně nazývá *tapping*. Důležitým

krokem je opravování pozic několikanásobným kritickým poslechem. U takto upravených anotací je předpokládána menší odchylka od reálných pozic dob. Uvedený způsob značení byl využit např. v rámci *Mazurka project*. [13]

Anotace velkého počtu interpretací je časově velmi náročný proces. Jednou z možností, jak zautomatizovat tento proces, je využít pokročilé metody beat trackingu. Možným efektivnějším způsobem je automatické převedení anotace jedné nahrávky na další nahrávky, a to např. využitím audio-to-audio synchronizace za pomoci metody DTW. Cílem takové synchronizace je stanovit zarovnání mezi hudebně odpovídajícími časovými pozicemi. Výhodou postupu přenosu anotace je, že na všech nahrávkách bude správný počet pozic, což usnadní kontrolu a případnou manuální korekci.

Pro účely diplomové práce byla v prostředí Sonic Visualiser manuálně vytvořena referenční anotace za použití nahrávky verze PHQ (seznam všech nahrávek je uveden v příloze B). Nahrávka byla vybrána z důvodu dobré poslechové kvality a rozlišitelnosti jednotlivých hlasů. Údaje o pozicích dob byly exportovány do souboru formátu .csv. Následná synchronizace je popsána v kapitole 6.

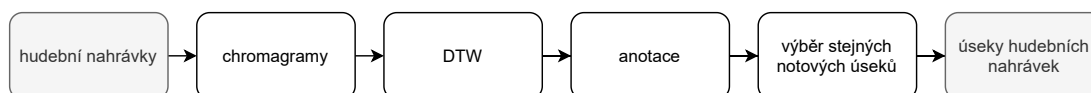
Na obr. 5.1 je zobrazen příklad anotace 6taktového hudebního úseku v prostředí Sonic Visualiser. Průběh signálu je znázorněn fialovou barvou, zelenou barvou jsou označeny pozice jednotlivých dob a oranžově jsou odděleny jednotlivé takty.



Obr. 5.1: Ukázka anotace úvodu 2. věty *Smyčcového kvartetu č. 12 F dur* v prostředí Sonic Visualiser.

## 6 Synchronizace nahrávek

Pro účely diplomové práce byla vytvořena databáze, která byla blíže popsána v kapitole 5. Cílem synchronizace není jen následné vybrání stejného notového úseku, ale také vytvoření anotací podle referenční anotace u všech nahrávek. Výsledkem je možnost vybrat si požadovaný hudební úsek podle čísla taktu, nebo i konkrétní doby. Na obr. 6.1 je zobrazen proces synchronizace nahrávek za pomoci algoritmu dynamického borcení časové osy (DTW) a vybrání stejného hudebního úseku.

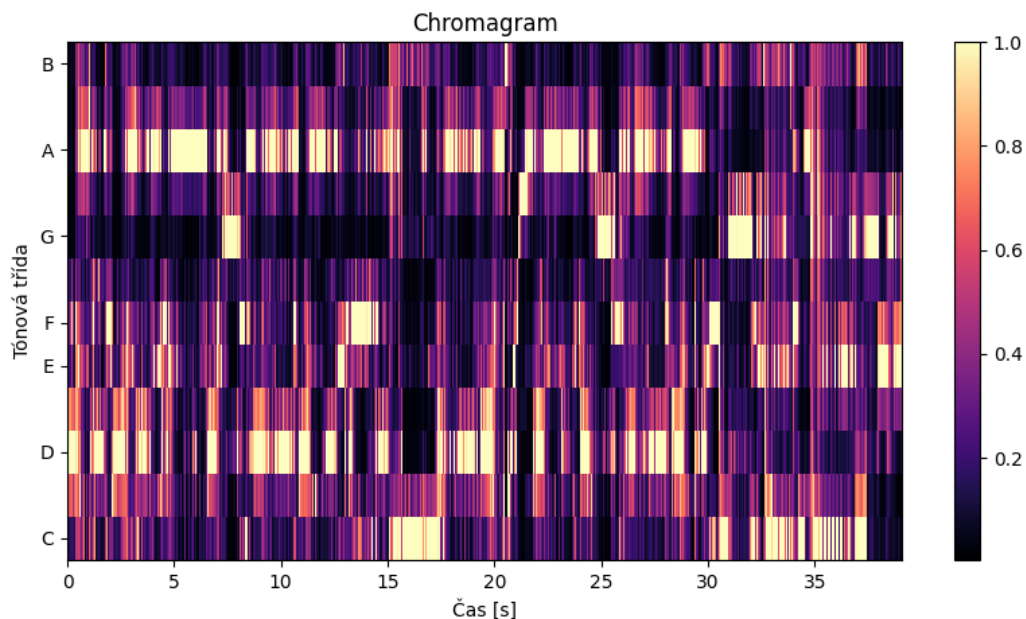


Obr. 6.1: Blokové schéma procesu synchronizace nahrávek a výběru stejného hudebního úseku.

### 6.1 Chroma reprezentace

V prvním kroku synchronizace jsou data z audio nahrávek převedena na chromagramy, sekvence chroma vektorů. Blíže byly chroma parametry rozebrány v kapitole 3.4.

Chromagramy charakterizují tónový obsah v nahrávce. Nahrávky rozdílných interpretací stejné skladby se vyznačují rozdílným tempem a dynamikou, společným rysem jim ale zůstává téměř stejný vývoj tonálního obsahu. Důvodem, proč jsou chroma parametry vhodné pro synchronizaci, je právě jejich závislost na tónových třídách. Veškerá spektrální energie signálu je v tomto procesu koncentrována do jedné oktávy a parametry jsou tak schopny spolehlivě zachytit melodickou strukturu i při proměnlivé barvě zvuku.



Obr. 6.2: Chromagram segmentu nahrávky realizované souborem Smetanovo kvarteto (1959).

## 6.2 Dynamické borcení časové osy

Dynamické borcení časové osy (Dynamic Time Warping, DTW) je metoda pro synchronizaci dvou posloupností podobného obsahu, ale rozdílné délky. Pro účely MPA se jedná např. o audio-to-score synchronizaci (audio signál vůči notové reprezentaci), nebo o audio-to-audio synchronizaci (dvě nahrávky se stejným obsahem, ale nejsou zarovnány na časové ose). V případě této práce se jedná o synchronizaci chromagramů hudebních nahrávek. Metoda DTW je založena na algoritmu využívajícího principu dynamického programování<sup>1</sup>.

Dva signály lze popsat jako posloupnosti:

$$X = x_1, x_2, \dots, x_i, \quad Y = y_1, y_2, \dots, y_j. \quad (6.1)$$

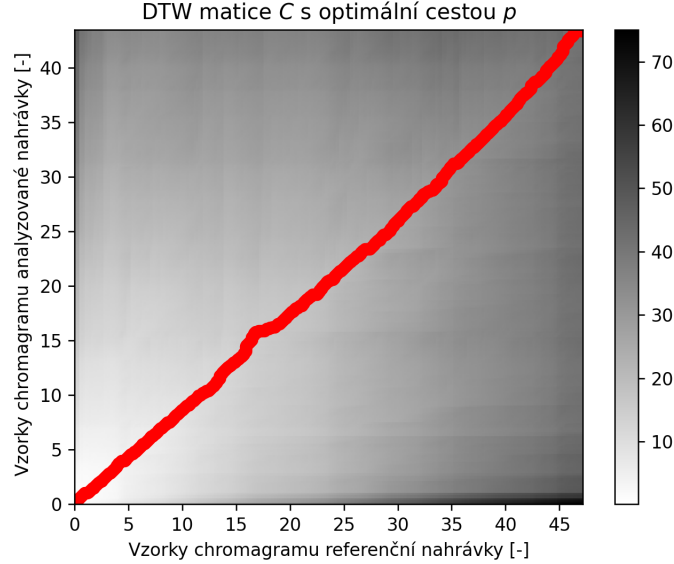
Základní myšlenkou metody DTW je pak nalézt optimální cestu dvojic indexů (tzv. *warping path*) mezi těmito posloupnostmi

$$\min \sum_{(i,j) \in P} p(x[i], y[j]), \quad (6.2)$$

<sup>1</sup>Princip dynamického programování spočívá v rekurzivním dělení úlohy na menší části, které se řeší ve vhodném pořadí, jejich výsledky se zaznamenávají a jsou použity pro řešení složitějších podúloh včetně původní úlohy.[33]

$$P = p_1, p_2, \dots, p_k. \quad (6.3)$$

Optimální cesta  $P$  je vyjádřena jako matice o dvou sloupcích, kde první sloupec obsahuje prvky referenčního chromagramu a druhý sloupec prvky synchronizovaného chromagramu.



Obr. 6.3: Ukázka DTW matice  $C$  s optimální cestou  $P$ .

Metoda DTW je sice poměrně starou metodou, ale pro svou jednoduchost a praktičnost je i v současnosti stále často využívána. Obecně se jedná o metodu, která se snaží nalézt optimální shodu mezi dvěma danými posloupnostmi s určitými omezeními a pravidly:

- každá souřadnice z první posloupnosti musí být spárována s jednou nebo více souřadnicemi z posloupnosti druhé a naopak;
- první souřadnice z první posloupnosti musí být spárována s první souřadnicí z druhé posloupnosti (nemusí to být však její jediná shoda)

$$p_1 = (x_1, y_1); \quad (6.4)$$

- poslední souřadnice z první posloupnosti musí být spárována s poslední souřadnicí z druhé posloupnosti (nemusí to být však její jediná shoda)

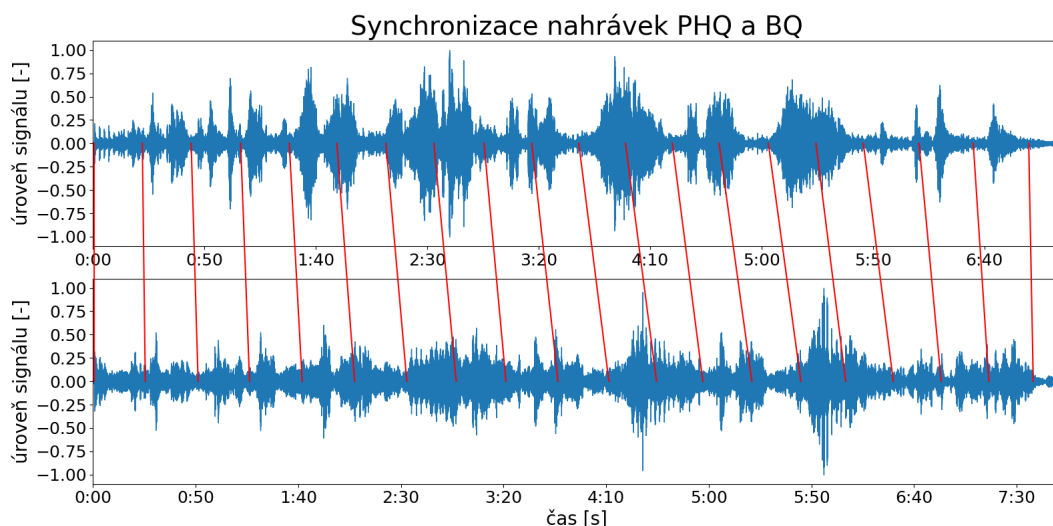
$$p_k = (x_i, y_j); \quad (6.5)$$

- Přiřazování souřadnic z první posloupnosti k indexům posloupnosti druhé musí postupně stoupat, a to platí i naopak. Validními kroky jsou tedy  $(i + 1, j)$ ,

$(i, j + 1)$  nebo  $(i + 1, j + 1)$ .

## 6.3 Vytvoření anotace

Pozice dob jednotlivých nahrávek jsou zarovnávány podle referenční anotace manuálně vytvořené u nahrávky uskupení Pavel Haas Quartet. Na obr. 6.4 je uveden příklad synchronizace. Pro přehlednost je vyobrazena pouze 1. doba každého 5. taktu.



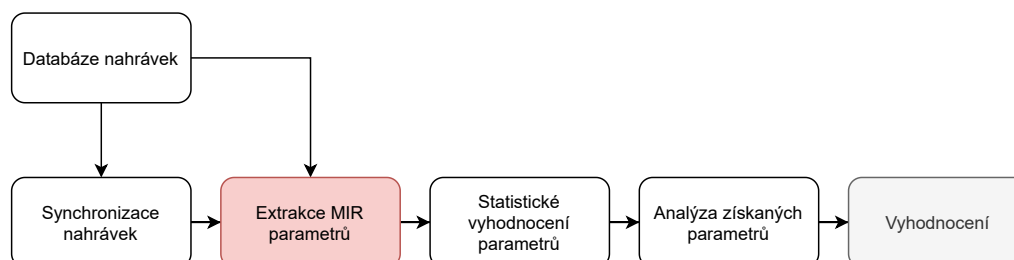
Obr. 6.4: Ukázka synchronizace pozic taktů nahrávek Pavel Haas Quartet a Borusan Quartet.

Nástroj je implementován v programovacím jazyce Python za použití balíčku `librosa`. Skript je navržen tak, že podle referenční nahrávky a její anotace postupně vytváří anotace ostatních nahrávek z databáze. Synchronizace je provedena za pomoci funkce `librosa.sequence.DTW`, která zachovává postup podle *Sakoe-Chiba* DTW algoritmu. Na závěr jsou jednotlivé anotace uloženy ve formátu `.csv`.

Výsledné vytvořené anotace jsou přesné na úrovni taktů, ne tolik už na úrovni jednotlivých dob. V závěru procesu je nutné provést manuální korekce jednotlivých pozic dob na základě kritického poslechu. Výsledkem je mimo jiné možnost vybrat hudební úsek podle taktu nebo doby.

## 7 Návrh vyhodnocovacího systému

Vyhodnocování odlišností hudebních nahrávek z databáze probíhá na základě vyhodnocovacího systému, jehož návrh je schematicky zobrazen na obr. 7.1.



Obr. 7.1: Návrh vyhodnocovacího systému.

K provedení experimentu je v prvním kroku třeba vytvořit databázi několika nahrávek stejné skladby od různých interpretů, s různým datem vzniku a různým původem interpretů. Databáze vytvořená pro tuto práci je podrobněji popsána v kapitole 5. V rámci synchronizace signálu jsou vytvořeny anotace pro všechny interpretace a vybírán stejný hudební úsek.

V dalším kroku jsou z hudebních stop extrahovány MIR parametry. Výpočty jsou uskutečněny v programovacím jazyce Python. Ze získaných dat jsou vypočteny statistické parametry. Výstupy jsou v podobě tabulek hodnot a grafů.

V následujícím kroku je provedena analýza získaných parametrů. Interpretace jsou mezi sebou porovnány z hlediska agogiky. V rámci statistické analýzy jsou roztrženy do tříd podle algoritmu Affinity Propagation (metoda strojového učení bez učitele).

## 8 Analýza získaných parametrů

Dříve v kapitole 1 byly rozebrány aspekty analýzy interpretačního výkonu. Při srovnávání charakteristik interpretací je na prvním místě vhodné uvažovat tempo a dynamiku. Jsou to základní parametry interpretace, se kterými interpret může pracovat, mimo jiné tyto parametry zrcadlí i poučenost a zkušenost hráče. Dalším hlediskem, na které je soustředěna čím dál větší pozornost, je míra expresivity. Obecně lze míru expresivity přednesu ohraničit označeními, že se jedná o velmi expresivní interpretaci nebo naopak všední či „strojovou“ interpretaci. Jde však o komplexní problém. Dalo by se tvrdit, že např. expresivní interpretace jsou charakteristické častým *vibratem*. Zároveň však může být *vibrato* znakem nervozity performerů.

Vytvořená databáze obsahuje nahrávky českých a zahraničních smyčcových kvartet. Nejedená se však o rovnocenné zastoupení, pouze 12 nahrávek je českého původu. U většiny nahrávek je znám rok pořízení nahrávky, u 5 z nich se dataci bohužel nepodařilo dohledat. Polovina nahrávek byla pořízena až po roce 2000. Podařilo se získat pouze 4 nahrávky vzniklé před rokem 1970.

Postup analýzy byl zvolen následovně:

- porovnat lokální a globální tempo,
- metodou shlukovacího algoritmu získat podle extrahovaných parametrů skupiny nahrávek, které může spojovat určitý znak,
- hledat společné rysy mezi možnými skupinami nahrávek, např. mezi nahrávkami českého a zahraničního původu.

V tabulce 8.1 je uveden výčet extrahovaných parametrů:

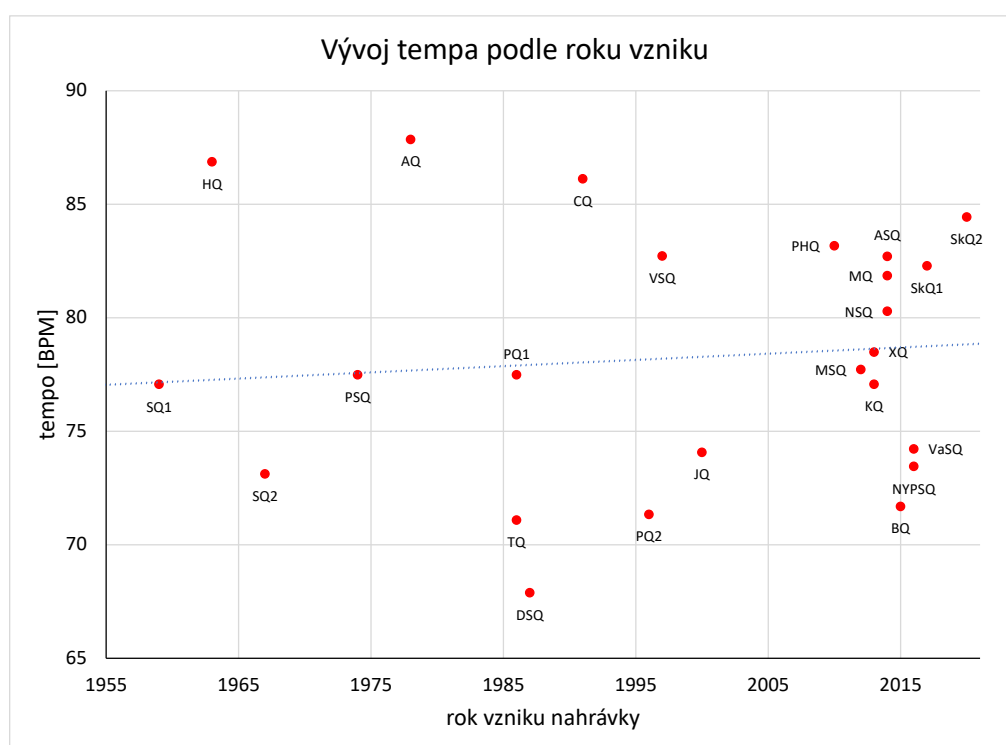
Tab. 8.1: Přehled extrahovaných parametrů

Název parametru	Zvolená zkratka
Efektivní hodnota signálu	<b>rms</b>
Tempo	<b>tempo</b>
Počet průchodů nulovou úrovní	<b>zero</b>
Spektrální centroid	<b>sp_cent</b>
Spektrální šířka pásma	<b>sp_band</b>
Ujíždění spektra	<b>sp_roll</b>
Melovské spektrální koeficienty (2–13)	<b>mfcc02–13</b>



## 8.1 Porovnání interpretací z hlediska agogiky

V prvním případě bylo porovnáno základní (průměrné) tempo každé nahrávky. Na obr. 8.1 lze pozorovat rozmístění průměrného tempa jednotlivých interpretací v závislosti na roku vzniku. Tempové označení *lento* podle rozdílné literatury se pohybuje v rozsahu 45–68 BPM. Z grafu lze vidět, že pouze interpret Delmé String Quartet (DSQ) zvolil tempo pomalejší než 70 BPM (konkrétně 67,9 BPM), tedy dle předpisu. Většina interpretů zvolila tempo mezi 75–85 BPM. Průměrná hodnota BPM pro nahrávky zahraničního původu je 77,8 BPM, pro české nahrávky je to 78,2 BPM.

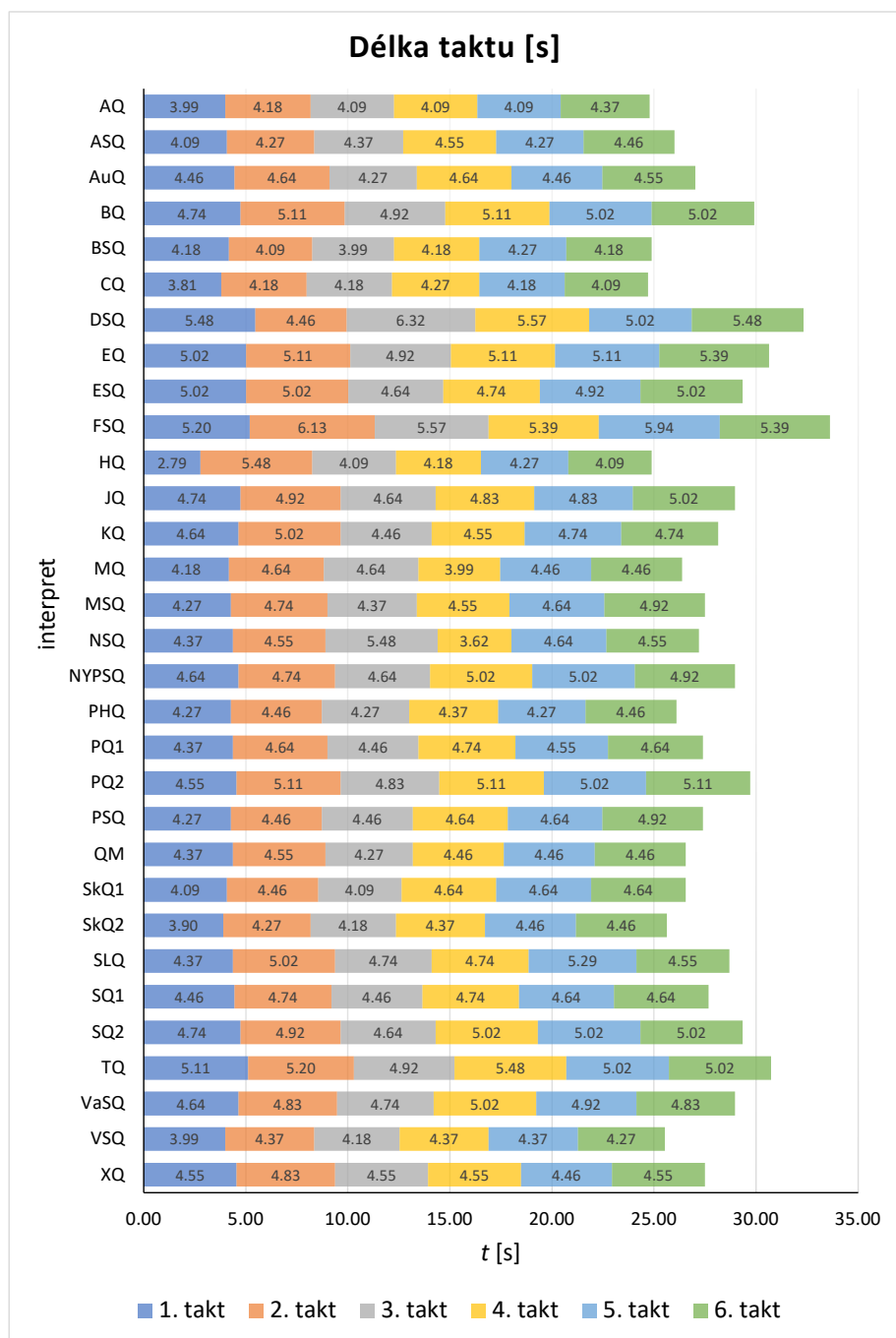


Obr. 8.1: Závislost hodnot průměrného tempa podle roku vzniku nahrávky.

Na obr. 8.3 jsou zobrazeny délky taktů 1–6, které naznačují modifikace vnitřní struktury. Pro přehlednost je přiložen notový zápis 1. a 2. houslí vybraného úseku na obr. 8.2. Obecně lze pozorovat trend, kdy je tempo zpomalené před nástupem 1. houslí ve 3. taktu (tzn. čas taktu se prodlouží). Následně je tempo určeno vedením hlasu 1. houslí, ve 3. taktu se vrátí přibližně k počátečnímu tempu a v závěru fráze v 6. taktu opět dojde k výraznému zvolnění. U některých interpretů např. u XIAO 4 Quartet (XQ), Smetanova kvarteta (verze z roku 1959, SQ1) nebo u Borromeo String Quartet (BSQ) je vidět stabilní průběh tempa. Oproti tomu lze pozorovat sklony ke zpomalování a zrychlování např. u tělesa Delmé String Quartet (DSQ) nebo u Panochova kvarteta (verze z roku 1996, PQ2).

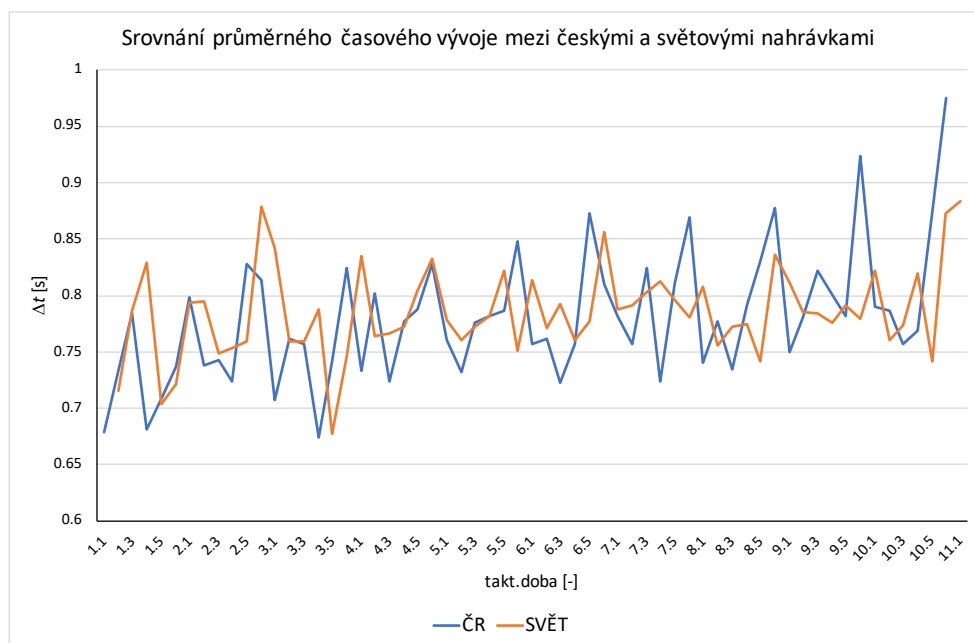


Obr. 8.2: Notová transkripce taktu 1–6, 1. a 2. housle.



Obr. 8.3: Srovnání délky taktů 1–6 všech interpretací z databáze.

Na obr. 8.4 je zobrazena posloupnost časových rozdílů mezi jednotlivými dobami první fráze skladby (takt 1–10). Modrá křivka reprezentuje nahrávky českého původu, oranžová křivka nahrávky světového původu. V obou případech lze sledovat trend „houpavého“ tempa, od 1. do 4. doby tempo lehce zrychluje, na 5. a 6. době zase zpomaluje. U zahraničních nahrávek lze pozorovat mírný rozdíl ve 2. polovině fráze, od 7. taktu dochází k postupnému zvolňování.



Obr. 8.4: Srovnání časového vývoje průměru českých a zahraničních nahrávek.

Pro vizualizaci podobnosti (či odlišnosti) interpretací mezi sebou lze využít korelační grafy. V rámci projektu *Mazurka Project*, zmíněného v kapitole 1.1, byl vytvořen volně dostupný nástroj pro tvorbu těchto grafů. [14]

Korelace je metodou, která umožňuje charakterizovat podobnosti dvou posloupností. Obecně se jedná o sumu součinů mezi odpovídajícími prvky dvou posloupností. Výsledek je normován na rozsah od  $-1$  do  $+1$  za pomoci Pearsonova korelačního koeficientu. Při korelaci lze vyjádřit pouze celkovou podobnost, jedním výsledným číslem však nelze charakterizovat podobnost posloupností nějak podrobněji. Pro sledování vnitřních podobností je možné posloupnost rozdělit do menších úseků a hierarchicky postupovat od nejmenších bloků k celku. Tím jsou zavedeny hierarchické korelační grafy.

Na obr. 8.5 je znázorněna škála barevnosti využitá pro reprezentaci míry korelace<sup>1</sup>.

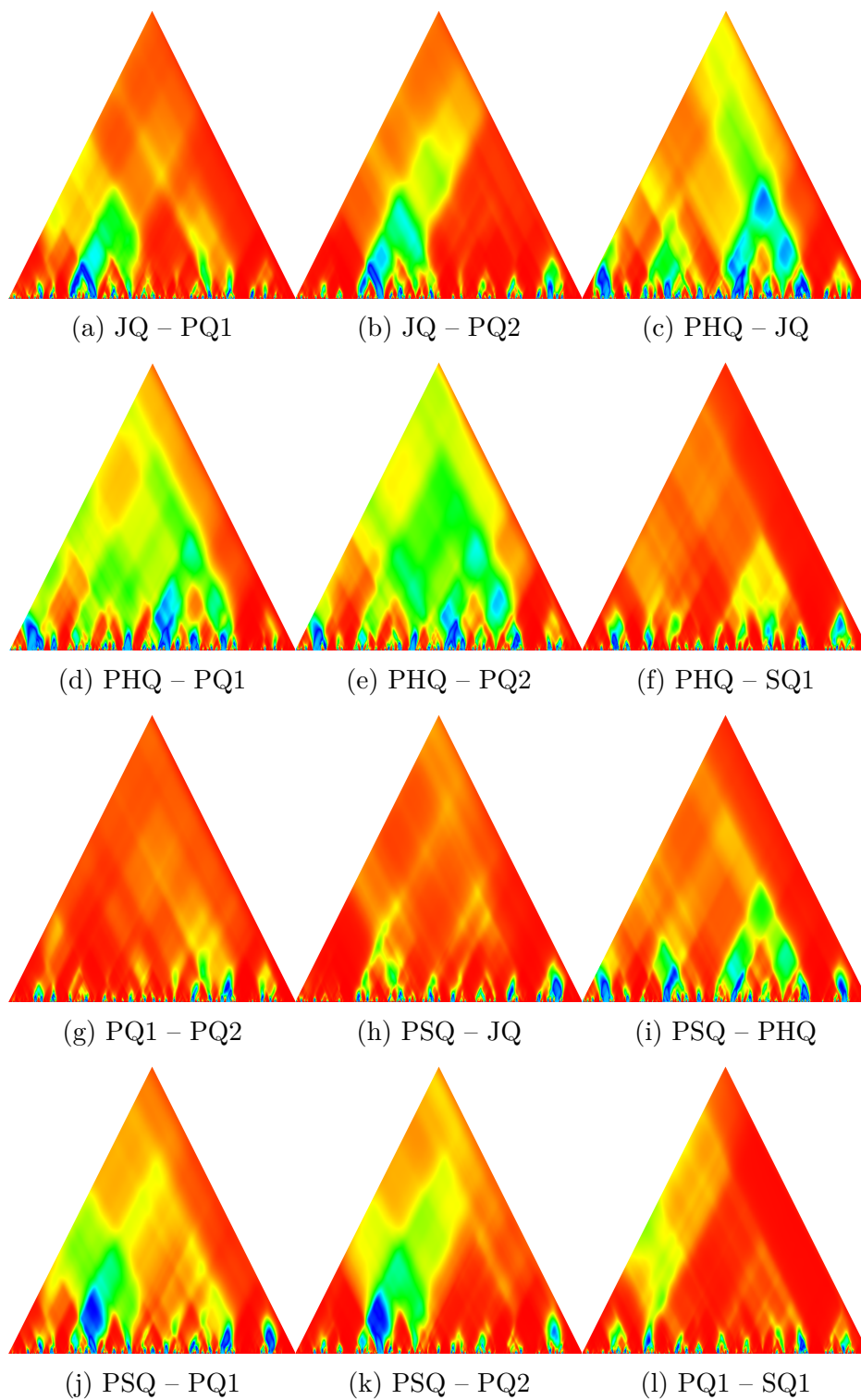


Obr. 8.5: Škála barevnosti korelačního grafu (červená = silně pozitivní korelace, fialová = silně negativní korelace).

Na obr. 8.6 je zobrazen výčet korelací průběhu tempa nahrávek vybraných českých interpretů. Ve spodní části grafu lze pozorovat podobnost vývoje tempa v čase na elementární úrovni dob. Směrem nahoru se porovnávané bloky zvětšují a na úplném vrcholu lze vyčíst celkovou podobnost tempa vybraných interpretací. V databázi jsou k dispozici dvě verze nahrávek různé datace u interpretů Panochovo kvarteto (PQ1 – PQ2), Smetanovo kvarteto (SQ1 – SQ2) a Škampovo kvarteto (SkQ1 – SkQ2). V případě Panochova kvarteta lze soudit, že nahrávky jsou si velmi podobné a interpreti zachovávají charakteristický projev, u Smetanova kvarteta lze vidět drobné rozdíly v úvodu skladby. Při posuzování vztahu mezi nahrávkou Kvarteta Pavla Haase a Panochova kvarteta (PHQ – PQ2) lze vidět jen velmi slabou pozitivní korelaci. Z grafu také vyplývá, že ve střední části si nahrávky nejsou podobné. Obdobně je tomu v případě Kvarteta Pavla Haase a Janáčkova kvarteta (PHQ – JQ). Když potom srovnáme interpretaci Janáčkova kvarteta a Panochova kvarteta (JQ – PQ2), tak u těchto nahrávek lze vidět silnější míru pozitivní korelace.

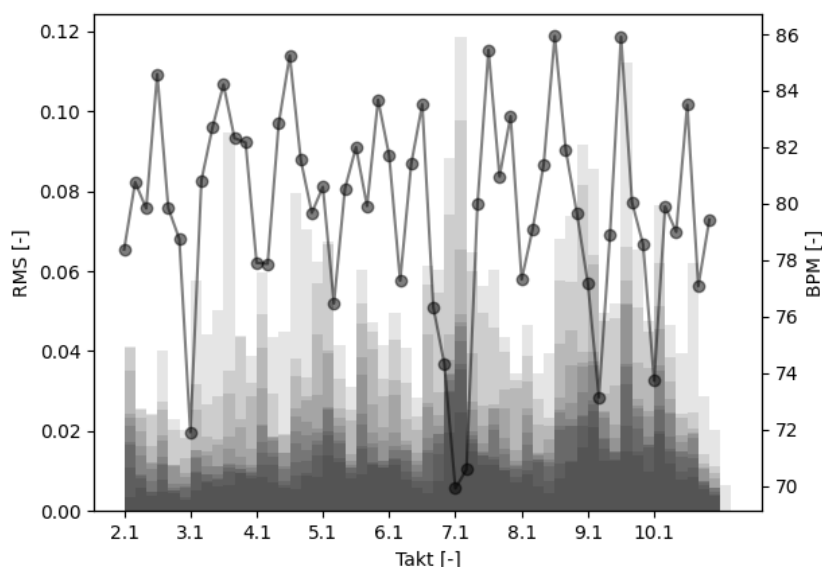
---

<sup>1</sup>*Hierarchical Correlation Plots*. Dostupné z URL: <<http://www.mazurka.org.uk/ana/timescape/>>



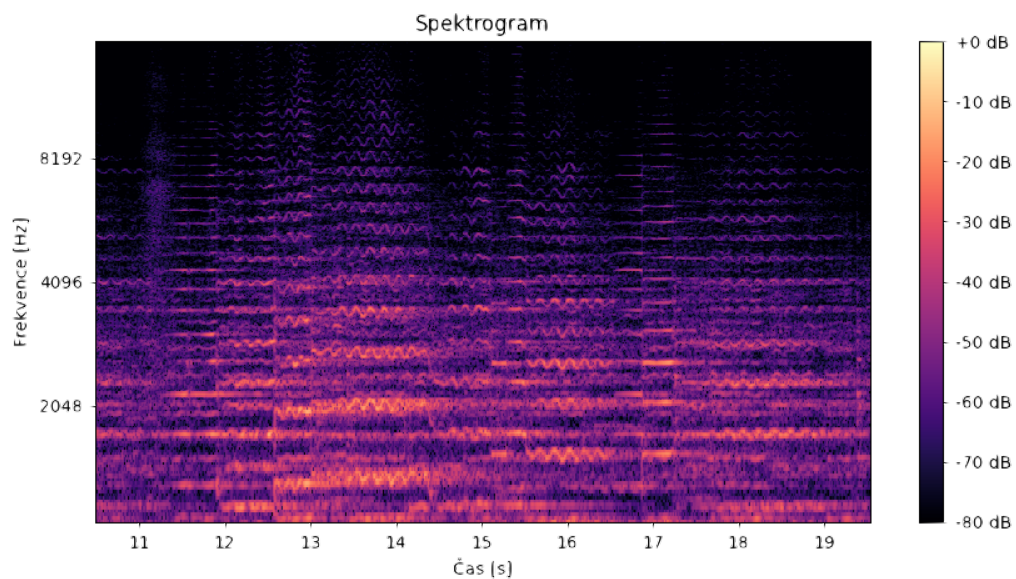
Obr. 8.6: Korelace vývoje tempa nahrávek vybraných českých interpretů.

Pro extrakci parametru RMS byla použita funkce `librosa.feature.rms`. Lze pozorovat návaznost vývoje RMS na vývoj tempa. U všech nahrávek přetrvává trend zeslabování při snižování tempa ke konci taktů, typicky na 5. a 6. době. Tuto návaznost lze sledovat v taktech 2–10 na obr. 8.7, kde spojnicový graf reprezentuje průměrné tempo nahrávek a sloupcový graf vývoj dynamiky.

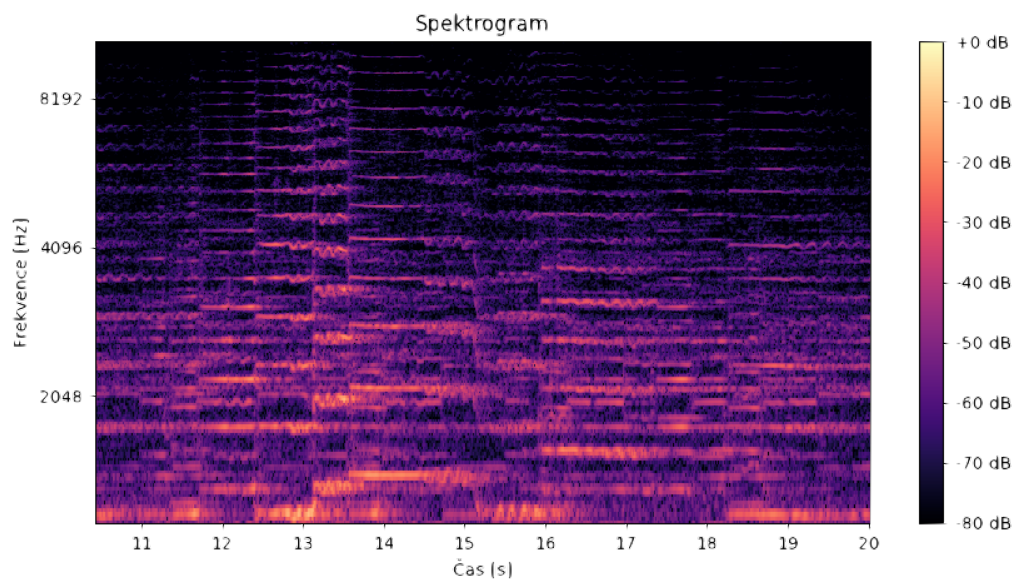


Obr. 8.7: Porovnání vývoje RMS s průměrným průběhem BPM.

Lze tvrdit, že v nahrávkách s většími výkyvy dynamiky a tempa je přítomna např. výrazně vyšší míra *vibrata*, což je samozřejmě typickým znakem expresivního projevu. Ukázka porovnání míry *vibrata* rozdílných interpretací je uvedena na obr. 8.8 (expresivnější projev) a obr. 8.9 (mírnější projev).



Obr. 8.8: Míra *vibrata* u interpretace souborem Pavel Haas Quartet (2010).



Obr. 8.9: Míra *vibrata* u interpretace souborem Musopen String Quartet (2012).

## 8.2 Klasifikace

Pro účely roztřídění získaných dat byla navržena aplikace algoritmu Affinity Propagation (AP). Jedná se o shlukovací algoritmus patřící mezi metody strojového učení bez učitele. Mezi podobné metody patří např. shlukovací algoritmus  $k$ -means, kdy je při vstupu dat definován počet shluků ( $k$ ), a poté jsou data do těchto kategorií přiřazena podle vzdálenosti mezi prvky a těžištěm shluku. Ovšem zvolit správný počet kategorií může být problematické. Oproti tomu u algoritmu Affinity Propagation není nutné předem definovat počet kategorií. Cílem tedy bylo aplikovat algoritmus Affinity Propagation, získat shluky dat a hledat společné znaky, které mohou takto vytvořené shluky charakterizovat.

Algoritmus AP vyměňuje informace mezi daty dokud se neobjeví sada vzorů, kde každý vzor reprezentuje jeden shluk. Vstupem pro algoritmus jsou reálná čísla  $s(k, k)$  pro každý vzorek  $k$ , označovaná jako preference. Vzorky s vyššími hodnotami pro  $s(k, k)$  mají vyšší pravděpodobnost stát se vzory. Počet klastrů je ovlivněn hodnotami preferencí a průběhem předávání zpráv [8]. Shlukovací algoritmus AP je obsažen v knihovně pro strojové učení `scikit-learn`, v podobě funkce `sklearn.cluster.AffinityPropagation`. Algoritmus byl aplikován na parametry extrahované z taktů 1–10.

Na obr. 8.10 jsou zobrazeny získané shluky v případě závislosti proměnlivosti tempa a dynamiky. K tomu byly využity směrodatné odchylky hodnot tempa a RMS. Směrem k hodnotě  $-1$  se jedná o málo proměnlivé parametry, směrem k hodnotě  $+1$  se jedná o velmi proměnlivé parametry. Algoritmus AP vytvořil celkem 8 kategorií.

Při pozorování vytvořených kategorií lze vidět základní rozdělení nahrávek:

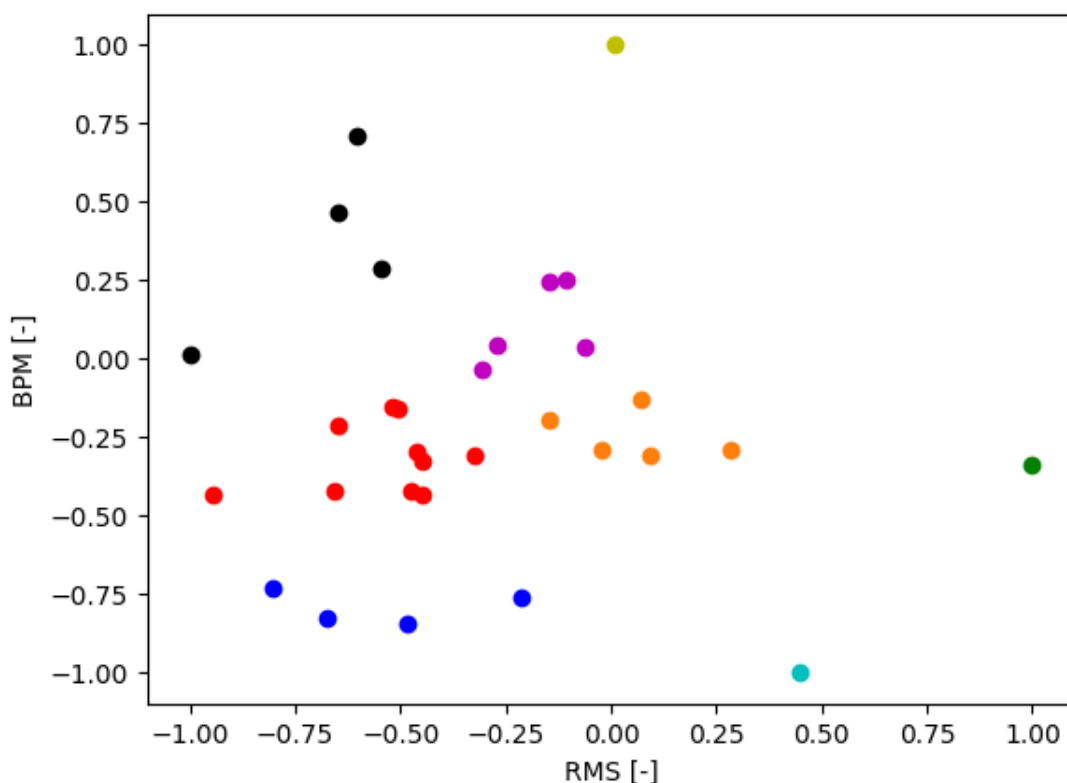
- nahrávky s pravidelným tempem a malými výkyvy dynamiky (modrá),
- nahrávky s pravidelným tempem a velkými výkyvy dynamiky (tyrkysová, zelená),
- nahrávky o něco hybnější s malými výkyvy dynamiky (červená),
- nahrávky s proměnlivým tempem a dynamikou (fialová a oranžová),
- nahrávky s proměnlivým tempem a malými výkyvy dynamiky (černá).

V případě fialového a oranžového shluku je možné mluvit o průměrné interpretaci. Patří sem interpreti ASQ, DSQ, NYPSQ, QM, CQ, KQ, PQ2, SkQ1, TQ. Jsou zde zastoupeni čeští i zahraniční interpreti.

V případě tělesa Novus String Quartet (zelená) se podle výsledku jedná o nahrávku s výrazně proměnlivou dynamikou a stabilnějším průběhem tempa. Jedná se o jediný prvek ve své kategorii. Je možné, že nahrávka byla výrazně ovlivněna v postprodukci úpravou dynamiky. Nahrávka s úzkým dynamickým profilem a mírně variabilním tempem je od uskupení Euclid Quartet (černá). Podle toho by se i další 3 nahrávky v kategorii měly vyznačovat podobnou proměnlivostí tempa a ne příliš



expresivní dynamikou. Další krajní polohu lze nalézt u interpretace od Smetanova kvarteta z roku 1959 (žlutá), v tomto případě se jedná o tempově proměnlivou nahrávku s výraznější dynamikou, opět se jedná o prvek jediný ve své kategorii. Při kontrolním poslechu je tempo proměnlivé nejen na úrovni taktů na 5. a 6. době (jak bylo zmíněno v kapitole 8.1), ale také v závislosti na vedoucích hlasech (ve frázi s vedením 1. houslí je rychlejší tempo než ve frázi s vedením violoncella). Poslední krajní polohu zastává interpretace Škampova kvarteta z roku 2020 (tyrkysová), která je jediným zástupcem ve vytvořené kategorii. Je zde očekáváno velmi stabilní tempo a širší dynamický profil. Pravidelnost tempa uvedené interpretace si lze ověřit i na dříve uvedeném grafu 8.3.

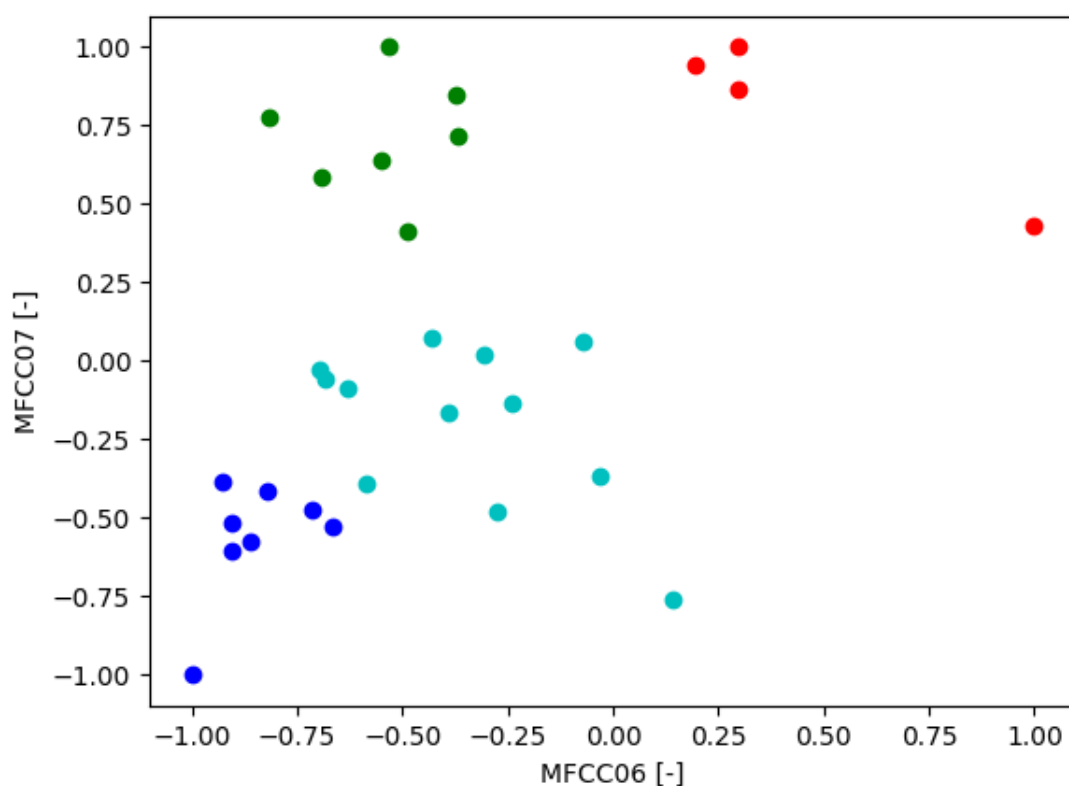


Obr. 8.10: Shluky při závislosti RMS a tempa.

Algoritmus AP byl aplikován na všechny kombinace extrahovaných parametrů. Následně bylo posuzováno, zda nahrávky vytvořily výrazné shluky. V případě parametrů **zero** nedošlo k zajímavým pozorováním. U parametrů **sp\_band** a **sp\_roll** byly výsledné shluky velmi podobné jako u **sp\_cent**, proto byly podrobněji posuzovány shluky se **sp\_cent**.

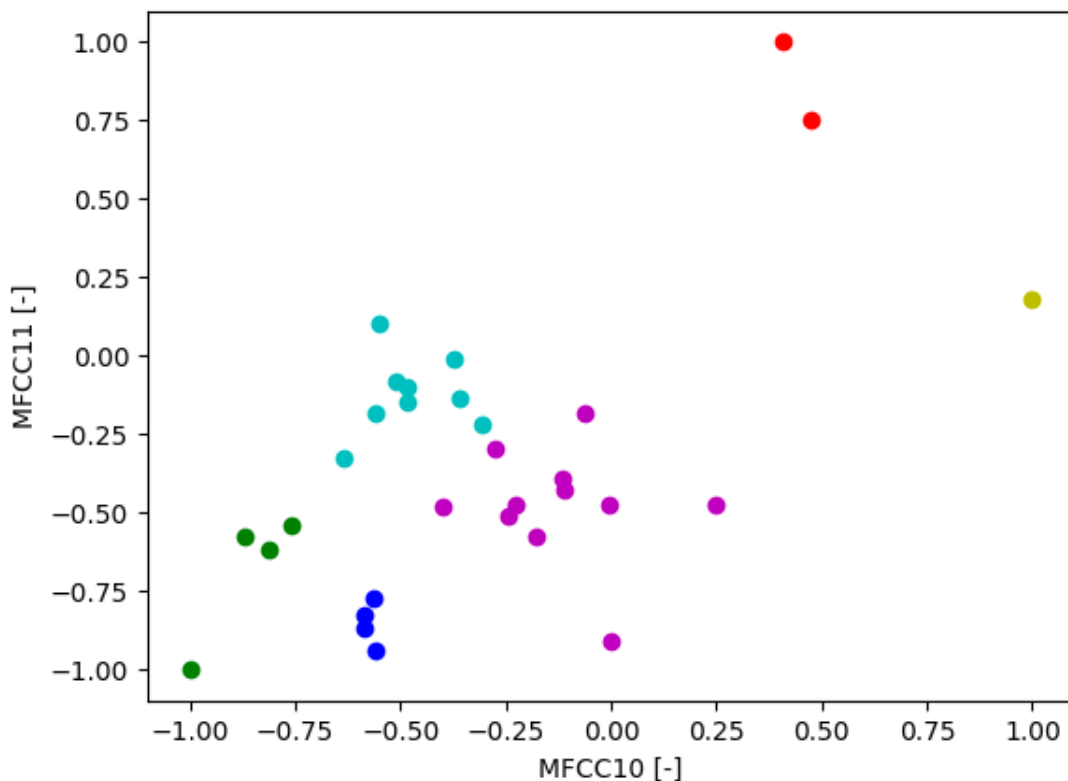
Na obr. 8.11 lze pozorovat nápadné shluky v souvislosti s 6. a 7. melovským

kepstrálním koeficientem. V případě tmavě a světle modré lze uvažovat jednu skupinu. Patří sem nahrávky AuQ, BQ, BSQ, CQ, EQ, ESQ, HQ, KQ, MSQ, NSQ, NYPSQ, PHQ, PQ2, SkQ1, TQ, VaSQ a XQ. Dvě verze interpretace Škampova kvarteta jsou od sebe relativně vzdálené, verze z roku 2017 náleží do světle modrého shluku, naopak jen o něco málo starší verze z roku 2020 náleží do červeného shluku. Ovlivněno to může být například zvolenou nahrávací místností a barvou dozvuku. Při kontrolním poslechu se např. červený shluk vyznačuje podobnou barvou 1. houslí a obecně zastoupením vyšších frekvencí ve spektru nahrávky. Prvky v zeleném shluku nespojuje žádný specifický rys.



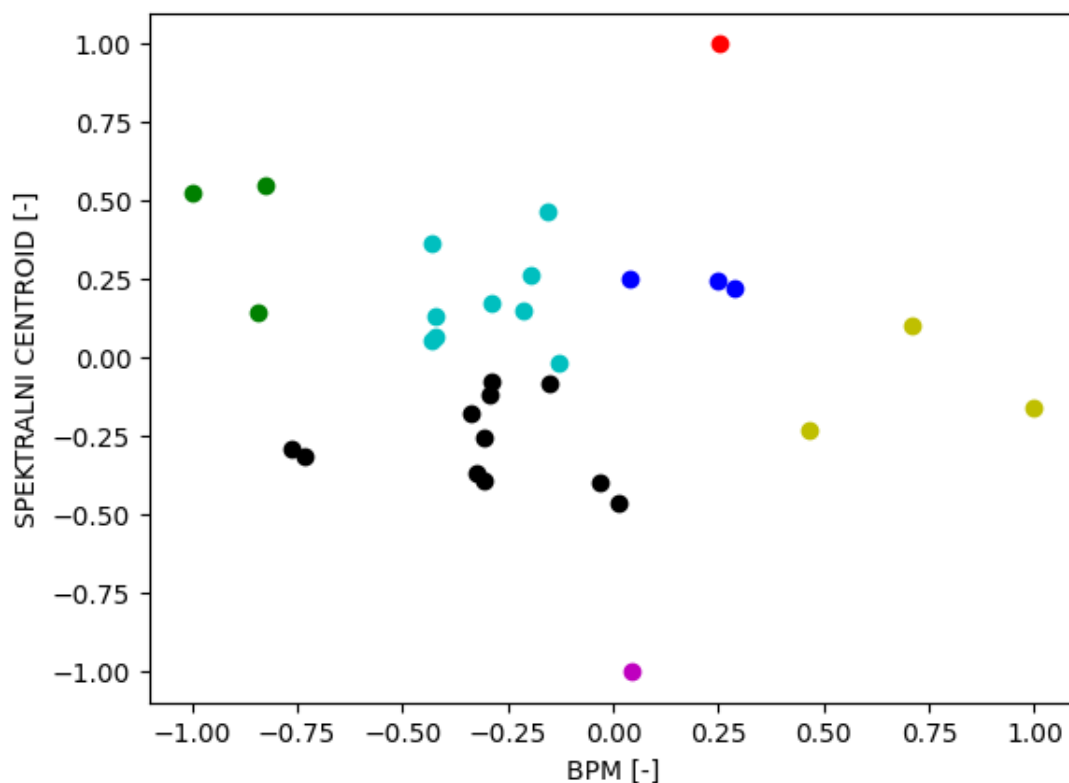
Obr. 8.11: Shluky při závislosti MFCCs 6 a 7.

Na obr. 8.12 je možné sledovat koncentraci shluků (zelená, tmavě modrá, světle modrá a fialová). Nápadně se vzdalují pouze 3 nahrávky, AQ a ASQ (červená) a SkQ2 (žlutá). Podobně jako na předchozím obr. 8.11 se jedná o nahrávky se světlejší barvou, která může být ovlivněna např. místem pořízení nahrávky.



Obr. 8.12: Shluky při závislosti MFCCs 10 a 11.

Na obr. 8.13 jsou vyobrazené shluky při závislosti tempa a spektrálního centroidu. Většina nahrávek je umístěna ve středu, jedná se o interprety AQ, ASQ, AuQ, BQ, BSQ, CQ, EQ, ESQ, FSQ, HQ, JQ, MQ, NSQ, NYPSQ, MSQ, PQ1, PQ2, SkQ1, TQ, VSQ a XQ. V krajní poloze se nachází uskupení Ševčíkovo-Lhotského kvarteto (fialová). Jedná se o jedinou velmi starou nahrávku v databázi, byla pořizena v roce 1929. Má tak výrazně jinou barvu než ostatní nahrávky. V případě Delmé String Quartet (červená) lze hovořit o jedné z nejpomalejších nahrávek z databáze. Ve žlutém shluku se nachází obě verze Smetanova kvarteta z let 1959 a 1967 a interpretace kanadského uskupení Vaughan String Quartet z roku 2016.



Obr. 8.13: Shluky při závislosti tempa a spektrálního centroidu.

Ve velkých shlucích se opakovaně objevují interpreti CQ, NYPSQ, PQ2, SkQ1, TQ a XQ. Nespojuje je ani rok vzniku ani původ, viz tabulka 8.2.

Výsledky ze shlukovací analýzy nejsou příliš vypovídající. V rámci práce nebylo po aplikaci shlukovacího algoritmu AP pro účely analýzy interpretačního výkonu dosaženo relevantních výsledků.

Tab. 8.2: Interpretace opakující se v hlavních shlucích

Interpret	Původ	Rok vzniku	Průměrné BPM
CQ	USA	1991	86,1
NYPSQ	USA	2016	73,4
PQ2	ČR	1996	71,3
SkQ1	ČR	2017	82,3
TQ	HUN	1986	71,1
XQ	USA	2013	78,5

# Závěr

Hudební interpretace je zásadním komunikačním článkem mezi autorem hudebního díla a posluchačem. Jedná se o fyzickou realizaci, kdy se každá autentická interpretace stává předmětem studia pro muzikologii, psychologii a v neposlední řadě pro oblasti zpracování signálu.

Analýza interpretačního výkonu se tradičně opírá o muzikologické a estetické teorie, nicméně s rozvojem výpočetních metod lze k analýze přistupovat čím dál více objektivněji. Stále prohlubující se poznatky v oblasti MIR (Music Information Retrieval – získávání informací o hudbě) umožňují analýzu automatizovat a nabízet nové nástroje pro výzkum v oblasti hudby.

Úvodní kapitoly práce mapují oblast analýzy interpretačního výkonu z muzikologického hlediska, která je nezbytná pro správné uchopení analýzy za pomoci metod MIR. Parametry, pomocí kterých je možné definovat interpretační výkon, lze rozdělit do následujících kategorií: tempo a čas, dynamika, témbra a výška tónu. Pro analýzu interpretačního výkonu byla navržena řada softwarových knihoven, např. `MIRtoolbox` pro MATLAB nebo `librosa` pro programovací jazyk Python. Významným projektem je interaktivní aplikace Sonic Visualiser, která umožňuje analýzu, vizualizaci a anotaci hudebních nahrávek.

Pro účely diplomové práce byla vytvořena databáze 31 nahrávek *2. věty Smyčcového kvartetu F dur, op. 12* od českého skladatele Antonína Dvořáka (1841–1904). V prostředí Sonic Visualiser byla zhotovena referenční anotace dob za použití nahrávky verze PHQ. Nahrávka byla vybrána z důvodu dobré poslechové kvality a rozlišitelnosti jednotlivých hlasů.

Pro zefektivnění procesu tvorby anotací byl navržen nástroj v programovacím jazyce Python, který by umožnil vytvořit anotace všech nahrávek z databáze podle referenční anotace. K tomu využívá algoritmu dynamického borcení časové osy. Výsledek je přesný na úrovni taktů, pozice jednotlivých bylo nutné v závěru procesu na základě kritického poslechu manuálně opravit.

Extrakce několika parametrů byla provedena v programovacím jazyce Python a v prostředí aplikace Sonic Visualiser. Vybrány byly základní nízkourovňové parametry – tempo, RMS, spektrální centroid, počet průchodů nulovou úrovní, chroma parametry, melovské keprální koeficienty. Následně byly výsledky posuzovány, jestli lze mezi nimi pozorovat určité trendy.

Rozdílné interpretace lze primárně odlišit individuálně voleným tempem a dynamikou. Pro vývoj energie v čase je potřeba sledovat parametr RMS, pro časové rozdíly je nutné sledovat tempo typicky vyjádřeného s jednotkou BPM. Spojení rozdílné artikulace a dynamiky tvoří jedinečnou barvu zvuku a tím jedinečnou interpretaci.

Při porovnávání tempa je možné sledovat podobný přístup k interpretaci u českých i zahraničních interpretů. Většina interpretů zvolila tempo mezi 75–85 BPM. Průměrná hodnota BPM pro nahrávky zahraničního původu je 77,8 BPM, pro české nahrávky je to 78,2 BPM. U všech nahrávek přetrvává trend zeslabování při snižování tempa ke konci taktů, typicky na 5. a 6. době.

Pro účely roztržení získaných dat byla navržena aplikace algoritmu Affinity Propagation. Cílem bylo aplikovat algoritmus pro získání shluků dat a hledat společné znaky, které mohou takto vytvořené shluky charakterizovat. Ve velkých shlucích se opakovaně objevují interpreti CQ, NYPSQ, PQ2, SkQ1, TQ a XQ. Nespojuje je ani rok vzniku, ani původ. V rámci práce nebylo po aplikaci shlukovacího algoritmu pro účely analýzy interpretačního výkonu dosaženo relevantních výsledků.

Jako možné zkreslení výsledků práce lze uvažovat rozdílnou kvalitu jednotlivých nahrávek, která může ovlivnit proces extrakce parametrů a následné vyhodnocení, nebo také např. problematiku ladění. Nástroje v jednotlivých nahrávkách nemusí být naladěny na stejnou frekvenci (tzv. *komorní a* odpovídá 440 Hz), a to může např. snížit úspěšnost algoritmu DTW.

V rámci diplomové práce byla provedena důkladná rešerše nejmodernějších přístupů k objektivní analýze interpretačního výkonu. Byly nastíněny aspekty interpretačního výkonu, které je možné podrobně porovnávat a zkoumat, a byly ukázány různé způsoby interpretace extrahovaných informací. Není snadné vystoupit z konvenčního muzikologického myšlení a hledat odlišné cesty analýzy než např. z hlediska tempového a dynamického vývoje. Současně není jednoduché a jasné vytvořit obecné kategorie ke klasifikaci. Příkladem může být snaha klasifikovat podle původu interpreta. Hudebníci a hudební tělesa se vždy vyznačovali migrací. I když je dostupný rok vzniku nahrávky a původ interpreta, i tak tyto informace mohou být vágní, tzn. interpret mohl studovat v jiné zemi a nahrát nahrávku zase v další zemi. Porovnávání parametry a příslušná metadata tedy nemusí nutně poskytovat validní informační hodnotu.

Současnou komplikací pro výzkum je vysoký nedostatek dostupných databází k testování a experimentování. Pro budoucí zlepšení podmínek výzkumu interpretačního výkonu je třeba soustředit pozornost na tvorbu databází anotovaných hudebních nahrávek odlišných interpretací, a to v různých kategoriích (sólový nebo sborový zpěv, sólové nástroje, komorní tělesa a symfonické orchestry).

# Literatura

- [1] FUKAČ, Jiří. *Agogika*. In *Slovník české hudební kultury*. Praha: Editio Supraphon, 1997. s. 22. ISBN 80-7058-462-9.
- [2] BÖCK, Sebastian a kol. *madmom: a new Python Audio and Music Signal Processing Library* [online]. 2016 [cit. 15. 5. 2021]. Dostupné z URL: <<https://arxiv.org/abs/1605.07008>>
- [3] BOŠTÍK, Milan. *Metody získávání budících pulsů* [online]. 2002 [cit. 15. 5. 2021]. Dostupné z URL: <<http://www.elektrorevue.cz/clanky/02038/index.html>>
- [4] BOWEN, José Antonio. *Tempo, Duration and Flexibility: Techniques in the analysis of performance*. In: *Journal of Musicological Research*. Vol. 16 [online]. 1996. s. 111–156 [cit. 21. 11. 2020]. Dostupné z URL: <<http://josebowen.com/wp-content/uploads/2013/06/Bowen.TempoDurationFlexibility1996.pdf>>
- [5] CAETANO, Marcelo. *Automatic Segmentation of the Temporal Evolution of Isolated Acoustic Musical Instruments Sounds Using Spectro-Temporal Cues*. In *13th International Conference On Digital Audio Effects, DASx 2010 Proceedings* [online]. 2010, s. 1–8 [cit. 5. 12. 2020]. Dostupné z URL: <<http://articles.ircam.fr/textes/Caetano10b/index.pdf>>
- [6] COOK, Nicholas. *Changing the musical object: Approaches to performance analysis*. In *In Music's Intellectual History: Founders, Followers and Fads* [online]. 2009 [cit. 17. 11. 2020]. Dostupné z URL: <<https://rilm.org/historiography/cook.pdf>>
- [7] CANNAM, Chris, Christian LANDONE a Mark SANDLER. *Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files*. In *Proceedings of the ACM Multimedia 2010 International Conference* [online]. [cit. 18. 11. 2020]. Dostupné z URL: <<https://sonicvisualiser.org/sv2010.pdf>>
- [8] *Clustering*. In *scikit-learn: User Guide* [online]. [cit. 20. 5. 2021]. Dostupné z URL: <<https://scikit-learn.org/stable/modules/clustering.html>>
- [9] DEVANEY, Johanna. *Automatically extracting performance data from recordings of trained singers*. In *Psychomusicology: Music, Mind and*

- Brain* [online]. 2011, č. 21(1–2), s. 108–136 [cit. 25.2.2021]. Dostupné z URL: <[https://www.academia.edu/1538110/Automatically\\_extracting\\_performance\\_data\\_from\\_recordings\\_of\\_trained\\_singers](https://www.academia.edu/1538110/Automatically_extracting_performance_data_from_recordings_of_trained_singers)>
- [10] DIMOV, Tomislav. *Short Historical Overview and Comparison of the Pitch Width and Speed Rates of the Vibrato Used in Sonatas and Partitas for Solo Violin by Johann Sebastian Bach as Found in Recordings of Famous Violinists of the Twentieth and the Twenty-First Centuries* [online]. [cit. 5.5.2021]. Dostupné z URL: <<https://search.proquest.com/openview/168bed7b3b33abf7a6b6efc9e4cfc6f6/1?pq-origsite=gscholar&cbl=18750&diss=y>>. Disertační práce. West Virginia University.
- [11] DVOŘÁKOVÁ-MAREŠOVÁ, Jiřina. *Metoda analýzy interpretačního výkonu*. In *Živá hudba*. [online]. 2012, č.3, s. 97–107 [cit. 5.12.2020]. Dostupné z URL: <[https://ziva-hudba.info/wp-content/uploads/pdf/141130131531\\_pdf\\_0.pdf](https://ziva-hudba.info/wp-content/uploads/pdf/141130131531_pdf_0.pdf)>
- [12] EREMENKO, Vsevolod a kol. *Performance Assessment Technologies for the Support of Musical Instrument Learning* [online]. [cit. 5.5.2021]. Dostupné z URL: <[https://repositori.upf.edu/bitstream/handle/10230/44130/eremenko\\_CSME2020\\_perf.pdf?sequence=1&isAllowed=y](https://repositori.upf.edu/bitstream/handle/10230/44130/eremenko_CSME2020_perf.pdf?sequence=1&isAllowed=y)>
- [13] GADERMAIER, Thassilo a Gergard WIDMER. *A Study of Annotation and Alignment Accuracy for Performance Comparison in Complex Orchestral Music*. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* [online]. [cit. 10.4.2021]. Dostupné z URL: <<https://archives.ismir.net/ismir2019/paper/000094.pdf>>
- [14] *Hierarchical Correlation Plots* [online]. [cit. 21.5.2021]. Dostupné z URL: <<http://www.mazurka.org.uk/ana/timescape/>>
- [15] JĘDRZEJEWSKA, Maria Klara a kol. *Generating musical expression of MIDI music with LSTM neural network* [online]. [cit. 10.4.2021]. Dostupné z URL: <<https://ieeexplore.ieee.org/document/8431033>>
- [16] KUNA, Milan, Miloš BLÁHA. *Čas a hudba. K dramaturgii časových prostředků v hudebně interpretačním výkonu*. Praha: Academia, nakladatelství Československé akademie věd, 1982.
- [17] LARTILLOT, Olivier. *MIRtoolbox 1.7: User-s Manual* [online]. Oslo: University of Oslo, 2017 [cit.10.11.2020]. Dostupné z URL: <<https://>>



- [//www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf](http://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf)>
- [18] LERCH, Alexander. *Software-Based Extraction of Objective Parameters from Music Performances* [online]. [cit. 1. 5. 2021]. Dostupné z URL: <[https://www2.ak.tu-berlin.de/~akgroup/ak\\_pub/2008/Lerch%202008\\_Software-Based%20Extraction%20of%20Objective%20Parameters%20from%20Music%20Performances.pdf](https://www2.ak.tu-berlin.de/~akgroup/ak_pub/2008/Lerch%202008_Software-Based%20Extraction%20of%20Objective%20Parameters%20from%20Music%20Performances.pdf)>. Disertační práce. Technická univerzita Berlín.
  - [19] LERCH, Alexander a kol. *An Interdisciplinary Review od Music Performance Analysis*. In *Transactions of the International Society for Music Information Retrieval* [online]. 2020 [cit. 1. 4. 2021]. Dostupné z URL: <<https://transactions.ismir.net/articles/10.5334/tismir.53/>>
  - [20] *Librosa documentation* [online]. [cit. 21. 11. 2020]. Dostupné z URL: <<https://librosa.org/doc/latest/index.html>>
  - [21] MAEZAWA, Akira, Kazuhiko YAMAMOTO a Takuya FUJISHIMA. *Rendering music performance with interpretation variations using conditional variational RNN*. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* [online]. [cit. 10. 4. 2021]. Dostupné z URL: <<https://archives.ismir.net/ismir2019/paper/000105.pdf>>
  - [22] MATOUŠEK, Vojtěch. *Automatizovaný přepis hudební skladby do notového zápisu* [online]. Brno, 2018 [cit. 18. 5. 2021]. Dostupné z URL: <[https://www.vutbr.cz/www\\_base/zav\\_prace\\_soubor\\_verejne.php?file\\_id=171538](https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=171538)>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce Ing. Petr Petyovský, Ph.D.
  - [23] MCFEE, Brian a kol. *librosa: Audio and Music Signal Analysis in Python*. In *Proceedings of the 14th Python in Science Conference (SciPy 2015)* [online]. [cit. 21. 11. 2020]. Dostupné z URL: <[https://brianmcfree.net/papers/scipy2015\\_librosa.pdf](https://brianmcfree.net/papers/scipy2015_librosa.pdf)>
  - [24] MÜLLER, Meinard a kol. *Signal Processing for Music Analysis*. In *IEEE Journal of selected topics in signal processing* [online]. [cit. 28. 11. 2020]. Dostupné z URL: <<https://www.ee.columbia.edu/~dpwe/pubs/MuEKR11-spmus.pdf>>
  - [25] ORNOY, Eitan, Shai COHEN. *Analysis of contemporary violin recordings of 19th century repertoire: Identifying trends and impacts*. In *Frontiers in*



- [36] ZENKL, Luděk. *ABC hudební nauky*. Praha: Editio Bärenreiter, 2003. ISBN 80-86385-21-3.

# Seznam symbolů a zkratek

<b>AMPACT</b>	The Automated Music Performance Analysis and Comparison Toolkit
<b>AP</b>	Affinity Propagation
$B$	počet úderů
<b>BPM</b>	Beats Per Minute – počet úderů za minutu
<b>CHARM</b>	Centre for History and Analysis of Recorded Music – Centrum pro historii a analýzu hudebních nahrávek
<b>CQT</b>	Constant-Q Transform – konstantní Q transformace
<b>DSP</b>	Digital Signal Processing – číslicové zpracování signálů
<b>DTW</b>	Dynamic Time Warping – borcení časové osy
$f$	kmitočet
$f_{vz}$	vzorkovací kmitočet
$I$	intenzita zvuku
$L_I$	hladina intenzity zvuku
<b>ISMIR</b>	International Symposium on Music Information Retrieval – Mezinárodní sympozium k získávání informací o hudbě
<b>LSTM</b>	Long Short-Term Memory – dlouho-krátkodobá paměť
<b>MFCCs</b>	Mel-Frequency Cepstral coefficients – melovské keprální koeficienty
<b>MIR</b>	Music Information Retrieval – získávání informací o hudbě
<b>MPA</b>	Music Performance Analysis – analýza interpretačního výkonu
$P$	výkon
<b>RMS</b>	Root Mean Square – kvadratický průměr
$S$	plocha
<b>STFT</b>	The Short-Time Fourier Transform – krátkodobá Fourierova transformace
$T$	tempo

# Seznam příloh

A	Přiložené soubory	61
B	Databáze nahrávek k analýze	62
C	Notový zápis	63

## A Příložené soubory

K diplomové práci je přiložen soubor s názvem „dp\_archiv\_laborova.zip“, který obsahuje soubor s hlavním dokumentem ve formátu .pdf a 3 adresáře. Adresář „database“ obsahuje podadresář „annotations“ s vytvořenými anotacemi ve formátu .csv a soubor „database\_lento.xlsx“ s údaji o databázi ve formátu .xlsx. V adresáři „script“ se nachází veškeré použité skripty v programovacím jazyce Python pro synchronizaci nahrávek, extrakci parametrů a analýzu získaných výsledků. Dále adresář obsahuje podadresář „data“, kam se ukládají veškerá data (tj. zvukové nahrávky, anotace, extrahované parametry, výstupy), a podadresář „defs“, který obsahuje definice funkcí. Poslední adresář „tempo\_plots“ obsahuje tabulku s výpočty s hodnotami extrahovanými z programu Sonic Visualiser pro srovnání tempa jednotlivých nahrávek.

Obsah souboru „dp\_archiv\_laborova.zip“:

- DP\_LaborovaAnna.pdf
- database
  - annotations
  - database\_lento.xlsx
- script
  - data
  - defs
  - annotation.py
  - clustering\_multi.py
  - clustering.py
  - correlation.py
  - DTW\_plots.py
  - extraction.py
  - remix.py
  - sixBars.py
  - statistics.py
- tempo\_plots

## B Databáze nahrávek k analýze

Tab. B.1: Databáze nahrávek k analýze

Označení	Interpret	Původ	Rok vzniku
<b>AQ</b>	Amadeus Quartet	UK	1978
<b>ASQ</b>	Amphion String Quartet	USA	2014
<b>AuQ</b>	Auner Quartett	AUT	nezískán
<b>BSQ</b>	Borromeo String Quartet	USA	nezískán
<b>BQ</b>	Borusan Quartet	TUR	nezískán
<b>CQ</b>	Cleveland Quartet	USA	1991
<b>DSQ</b>	Delmé String Quartet	UK	1987
<b>ESQ</b>	Emerson String Quartet	USA	1986
<b>EQ</b>	Euclid Quartet	USA	nezískán
<b>FSQ</b>	Fry Street Quartet	USA	nezískán
<b>JQ</b>	Janáčkovovo kvarteto	CZ	2000
<b>KQ</b>	Košické kvarteto	SK	2013
<b>MQ</b>	Mucha Quartet	CZ	2014
<b>MSQ</b>	Musopen String Quartet	USA	2012
<b>NSQ</b>	Novus String Quartet	KOR	2014
<b>PQ1</b>	Panochovo kvarteto	CZ	1986
<b>PQ2</b>	Panochovo kvarteto	CZ	1996
<b>PHQ</b>	Pavel Haas Quartet	CZ	2010
<b>PSQ</b>	Prague String Quartet	CZ	1974
<b>QM</b>	Quatuor Mona	ESP	nezískán
<b>SQ1</b>	Smetanovo kvarteto	CZ	1959
<b>SQ2</b>	Smetanovo kvarteto	CZ	1967
<b>SLQ</b>	Ševčíkovo-Lhotského kvarteto	CZ	1929
<b>SkQ1</b>	Škampa Quartet	CZ	2017
<b>SkQ2</b>	Škampa Quartet	CZ	2020
<b>TQ</b>	Takács Quartet	HUN	1986
<b>HQ</b>	The Hungarian Quartet	HUN	1963
<b>NYPSQ</b>	The New York Philharmonic String Quartet	USA	2016
<b>VaSQ</b>	Vaughan String Quartet	CAN	2016
<b>VSQ</b>	Vlach String Quartet	CZ	1997
<b>XQ</b>	XIAO 4 Quartet	USA	2013

## C Notový zápis

Obr. C.1: A. Dvořák - *Smyčcový kvartet č. 12 „americký“*, F dur. 2. věta Lento, intro.

The image displays the musical score for the introduction of the second movement of Antonín Dvořák's String Quartet No. 12, 'American'. The score is written for four string instruments: Violin I, Violin II, Viola, and Violoncello. The key signature is F major (one flat) and the time signature is 6/8. The tempo is marked 'Lento'. The score is divided into three systems. The first system shows the initial measures, with the Violoncello part marked 'pizz.' and 'p'. The second system continues the introduction, featuring dynamic markings such as 'p', 'mp molto espressivo', and 'fz p'. The third system concludes the introduction, with dynamic markings like 'mf' and 'dim.'. The score includes various musical notations such as slurs, ties, and fingerings (e.g., 5, 3).